

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: June 18, 2004, 12:57:29 ; Search time 23 Seconds  
(without alignments)  
17.957 Million cell updates/sec

Title: US-10-655-201-2

Perfect score: 44

Sequence: 1 KPSPPEE 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 500 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	100.0	9	4	US-09-101-2726-77
2	44	100.0	10	1	US-07-603-675-1
3	44	100.0	11	1	US-07-603-675-8
4	44	100.0	21	4	US-09-101-2726-38
5	44	100.0	157	4	US-08-142-5908-25
6	44	100.0	200	4	US-09-101-2726-73
7	44	100.0	208	4	US-09-101-2726-98
8	44	100.0	365	1	US-08-093-741-83
9	44	100.0	365	1	US-08-120-012-83
10	44	100.0	393	2	US-08-560-098A-44
11	44	100.0	393	3	US-08-967-024C-24
12	44	100.0	393	3	US-08-967-024C-25
13	44	100.0	411	1	US-08-087-163-1
14	44	100.0	411	1	US-08-286-748B-18
15	44	100.0	411	1	US-08-153-799-18
16	44	100.0	411	2	US-08-560-098A-48
17	44	100.0	411	3	US-09-181-816-1
18	44	100.0	411	4	US-09-403-736-2
19	44	100.0	430	1	US-07-942-157A-3
20	44	100.0	430	6	5219569-2
21	44	100.0	431	4	US-09-101-2726-1
22	44	100.0	431	6	5188829-1
23	44	100.0	432	2	US-08-560-098A-47
24	36	81.8	352	3	US-09-122-079-1
25	36	81.8	352	4	US-09-875-423-1
26	36	81.8	362	4	US-09-875-114-1
27	36	81.8	418	2	US-08-257-963B-2

28	36	81.8	418	4	US-08-367-841A-2	Sequence 2, Appl
29	36	81.8	418	4	US-08-520-373D-2	Sequence 2, Appl
30	36	81.8	418	5	PCT-US95-07201-2	Sequence 1, Appl
31	35	79.5	19	1	US-08-291-349A-1	Sequence 1, Appl
32	35	79.5	19	1	US-07-990-266-1	Sequence 1, Appl
33	35	79.5	19	2	US-08-480-133A-1	Sequence 1, Appl
34	35	79.5	19	3	US-08-394-748A-9	Sequence 9, Appl
35	35	79.5	19	4	US-09-675-922-13	Sequence 13, Appl
36	35	79.5	19	5	PCT-US93-11781-1	Sequence 1, Appl
37	35	79.5	19	5	PCT-US95-02478-9	Sequence 9, Appl
38	35	79.5	150	4	US-09-252-991A-29265	Sequence 29265, A
39	35	79.5	271	3	US-08-536-891A-1	Sequence 1, Appl
40	35	79.5	271	3	US-08-933-100B-14	Sequence 14, Appl
41	35	79.5	271	4	US-09-366-009-1	Sequence 1, Appl
42	35	79.5	271	4	US-08-809-156B-1	Sequence 1, Appl
43	35	79.5	271	4	US-09-043-981-1	Sequence 1, Appl
44	35	79.5	266	2	US-08-836-854-4	Sequence 4, Appl
45	35	79.5	368	2	US-08-836-854-17	Sequence 17, Appl
46	35	79.5	457	2	US-08-836-854-16	Sequence 16, Appl
47	35	79.5	457	4	US-09-366-009-22	Sequence 22, Appl
48	35	79.5	457	4	US-08-809-156B-22	Sequence 22, Appl
49	35	79.5	547	4	US-09-366-009-13	Sequence 13, Appl
50	35	79.5	547	4	US-08-809-156B-13	Sequence 13, Appl
51	35	79.5	549	2	US-08-836-854-11	Sequence 11, Appl
52	35	79.5	549	4	US-09-366-009-23	Sequence 23, Appl
53	35	79.5	549	4	US-08-809-156B-23	Sequence 23, Appl
54	35	79.5	573	4	US-09-366-009-30	Sequence 30, Appl
55	35	79.5	573	4	US-08-809-156B-30	Sequence 30, Appl
56	35	79.5	574	2	US-08-836-854-21	Sequence 21, Appl
57	35	79.5	574	2	US-09-463-296-1	Sequence 1, Appl
58	35	79.5	574	4	US-09-366-009-24	Sequence 24, Appl
59	35	79.5	574	4	US-08-809-156B-24	Sequence 24, Appl
60	35	79.5	826	4	US-09-366-009-14	Sequence 14, Appl
61	35	79.5	826	4	US-08-809-156B-14	Sequence 14, Appl
62	35	79.5	896	4	US-09-585-858-17	Sequence 17, Appl
63	35	79.5	1240	2	US-08-680-326-37	Sequence 37, Appl
64	35	79.5	1240	4	US-09-904-065-4	Sequence 4, Appl
65	35	79.5	1240	4	US-09-904-065-15	Sequence 15, Appl
66	35	79.5	1876	4	US-09-418-710-71	Sequence 71, Appl
67	35	79.5	1878	4	US-09-418-710-13	Sequence 13, Appl
68	35	79.5	1911	1	US-08-348-006B-5	Sequence 5, Appl
69	35	79.5	1911	2	US-08-800-825A-5	Sequence 5, Appl
70	35	79.5	1911	5	US-09-158-657-5	Sequence 5, Appl
71	35	79.5	1911	5	PCT-US94-10166-5	Sequence 5, Appl
72	35	79.5	2231	1	US-08-153-799-16	Sequence 16, Appl
73	35	79.5	2324	1	US-08-283-857-1	Sequence 1, Appl
74	35	79.5	2324	5	PCT-US95-09819-1	Sequence 1, Appl
75	35	79.5	2324	6	5455158-1	Sequence 1, Appl
76	35	79.5	2387	2	US-09-016-366A-12	Sequence 12, Appl
77	35	79.5	2466	5	PCT-US93-12687-2	Sequence 2, Appl
78	35	79.5	2466	5	US-08-551-356-2	Sequence 2, Appl
79	34	77.3	15	3	US-08-167-867-11	Sequence 11, Appl
80	34	77.3	15	4	US-09-317-993-11	Sequence 11, Appl
81	34	77.3	49	6	5252477-4	Sequence 6, Appl
82	34	77.3	52	2	US-08-301-915-5	Sequence 5, Appl
83	34	77.3	52	3	US-08-524-694A-5	Sequence 5, Appl
84	34	77.3	52	3	US-09-117-217-10	Sequence 10, Appl
85	34	77.3	52	4	US-09-454-156A-5	Sequence 5, Appl
86	34	77.3	52	4	US-09-735-487-10	Sequence 10, Appl
87	34	77.3	114	4	US-09-489-039A-13622	Sequence 13622, A
88	34	77.3	146	4	US-09-252-991A-22791	Sequence 22791, A
89	34	77.3	163	3	US-09-117-217-8	Sequence 8, Appl
90	34	77.3	163	4	US-09-735-487-8	Sequence 8, Appl
91	34	77.3	172	2	US-08-117-981-6	Sequence 6, Appl
92	34	77.3	172	2	US-08-477-081-6	Sequence 6, Appl
93	34	77.3	172	5	PCT-US93-02142-6	Sequence 6, Appl
94	34	77.3	300	4	US-09-277-078-2	Sequence 2, Appl
95	34	77.3	336	2	US-08-474-379C-88	Sequence 88, Appl
96	34	77.3	382	4	US-09-533-029-92	Sequence 92, Appl
97	34	77.3	382	4	US-07-688-352C-40	Sequence 40, Appl
98	34	77.3	384	2	US-08-474-379C-40	Sequence 40, Appl
99	34	77.3	384	3	US-09-146-249A-40	Sequence 40, Appl
100	34	77.3	384	3	US-08-206-188B-40	Sequence 40, Appl

101	34	77.3	462	3	US-08-068-392-3	Sequence 3, Appl1	174	32	72.7	1294	4	US-09-400-348-3	Sequence 3, Appl1
102	34	77.3	462	3	US-08-396-988-3	Sequence 3, Appl1	175	32	72.7	1321	1	US-08-261-822A-3	Sequence 3, Appl1
103	34	77.3	500	1	US-08-375-510-1	Sequence 1, Appl1	176	32	72.7	1331	5	PCT-US95-07744A-3	Sequence 3, Appl1
104	34	77.3	500	2	US-08-487-657-1	Sequence 1, Appl1	177	32	72.7	1426	4	US-09-710-092-14	Sequence 14, Appl1
105	34	77.3	500	2	US-08-816-155B-45	Sequence 45, Appl1	178	32	72.7	2037	3	US-09-306-998-3	Sequence 3, Appl1
106	34	77.3	500	3	US-09-079-587-45	Sequence 45, Appl1	179	31	70.5	9	1	US-08-178-570-38	Sequence 38, Appl1
107	34	77.3	500	4	US-09-309-572-16	Sequence 16, Appl1	180	31	70.5	9	5	US-08-369-643-38	Sequence 38, Appl1
108	34	77.3	500	4	US-09-718-096-16	Sequence 16, Appl1	181	31	70.5	9	5	PCT-US95-00147-38	Sequence 38, Appl1
109	34	77.3	512	3	US-08-463-210-8	Sequence 8, Appl1	182	31	70.5	42	4	US-09-418-563-1	Sequence 1, Appl1
110	34	77.3	512	3	US-09-124-900-2	Sequence 2, Appl1	183	31	70.5	46	2	US-08-256-790-4	Sequence 4, Appl1
111	34	77.3	512	4	US-08-463-028-8	Sequence 8, Appl1	184	31	70.5	47	3	US-09-091-814-26	Sequence 26, Appl1
112	34	77.3	553	1	US-07-683-957B-3	Sequence 3, Appl1	185	31	70.5	47	4	US-09-205-258-507	Sequence 507, App
113	34	77.3	638	1	US-07-688-352C-22	Sequence 22, Appl1	186	31	70.5	47	4	US-09-205-258-513	Sequence 513, App
114	34	77.3	640	4	US-09-252-991A-19951	Sequence 19951, A	187	31	70.5	48	3	US-09-107-532A-3633	Sequence 35, Appl1
115	34	77.3	686	2	US-08-942-521B-9	Sequence 9, Appl1	188	31	70.5	50	2	US-08-256-790-6	Sequence 6, Appl1
116	34	77.3	734	3	US-09-146-249A-85	Sequence 85, Appl1	189	31	70.5	53	2	US-08-256-790-8	Sequence 8, Appl1
117	34	77.3	734	3	US-08-206-188B-85	Sequence 85, Appl1	190	31	70.5	53	2	US-08-256-790-10	Sequence 10, Appl1
118	34	77.3	885	2	US-08-577-192-33	Sequence 33, Appl1	191	31	70.5	59	4	US-09-621-976-7621	Sequence 7621, Ap
119	34	77.3	885	2	US-09-079-630-33	Sequence 33, Appl1	192	31	70.5	85	4	US-09-252-991A-26629	Sequence 26629, A
120	34	77.3	886	2	US-08-474-379C-65	Sequence 65, Appl1	193	31	70.5	88	4	US-09-107-532A-3633	Sequence 3633, Ap
121	34	77.3	886	3	US-09-146-249A-65	Sequence 65, Appl1	194	31	70.5	95	4	US-09-252-991A-31932	Sequence 31932, A
122	34	77.3	886	3	US-08-206-188B-65	Sequence 65, Appl1	195	31	70.5	97	4	US-09-252-991A-23889	Sequence 23889, A
123	34	77.3	888	2	US-08-474-379C-12	Sequence 12, Appl1	196	31	70.5	99	4	US-09-107-532A-3698	Sequence 3698, Ap
124	34	77.3	888	3	US-09-146-249A-12	Sequence 12, Appl1	197	31	70.5	101	4	US-09-252-991A-21911	Sequence 21911, A
125	34	77.3	898	3	US-08-206-188A-12	Sequence 12, Appl1	198	31	70.5	107	4	US-09-252-991A-30716	Sequence 30716, A
126	34	77.3	900	1	US-07-688-352C-12	Sequence 12, Appl1	199	31	70.5	127	4	US-09-252-991A-24186	Sequence 24186, A
127	33	75.0	417	4	US-09-252-991A-23757	Sequence 23757, A	200	31	70.5	130	4	US-09-252-991A-29337	Sequence 29337, A
128	33	75.0	624	4	US-09-071-035-244	Sequence 244, App	201	31	70.5	138	4	US-09-540-236-2988	Sequence 2928, Ap
129	33	75.0	1223	4	US-09-071-035-236	Sequence 236, App	202	31	70.5	139	4	US-09-216-393B-64	Sequence 64, Appl
130	33	75.0	1301	4	US-09-071-035-234	Sequence 234, App	203	31	70.5	139	4	US-09-198-452A-1197	Sequence 1197, Ap
131	33	75.0	1301	4	US-09-071-035-238	Sequence 238, App	204	31	70.5	141	4	US-09-252-991A-15681	Sequence 15681, A
132	33	75.0	1301	4	US-09-071-035-242	Sequence 242, App	205	31	70.5	155	3	US-08-716-190-8	Sequence 8, Appl1
133	33	75.0	1306	4	US-09-134-000C-6670	Sequence 6670, Ap	206	31	70.5	157	3	US-08-716-190-4	Sequence 4, Appl1
134	33	75.0	1308	4	US-09-134-000C-6588	Sequence 6588, Ap	207	31	70.5	161	2	US-08-683-2628-34	Sequence 34, Appl
135	32	72.7	113	4	US-09-452-239-16	Sequence 16, Appl	208	31	70.5	161	3	US-09-361-707-34	Sequence 34, Appl
136	32	72.7	138	4	US-09-148-545-222	Sequence 222, App	209	31	70.5	169	4	US-09-149-476-354	Sequence 354, App
137	32	72.7	139	4	US-09-148-545-160	Sequence 160, App	210	31	70.5	176	4	US-09-489-039A-11985	Sequence 11985, A
138	32	72.7	214	4	US-09-289-942A-6	Sequence 6, Appl1	211	31	70.5	179	4	US-09-252-991A-17121	Sequence 17121, A
139	32	72.7	243	1	US-07-869-933-29	Sequence 29, Appl1	212	31	70.5	202	4	US-09-252-991A-29946	Sequence 29946, A
140	32	72.7	243	1	US-07-869-933-33	Sequence 33, Appl1	213	31	70.5	222	4	US-09-227-357-161	Sequence 161, App
141	32	72.7	243	1	US-08-201-879A-4	Sequence 4, Appl1	214	31	70.5	255	3	US-09-040-485-2	Sequence 2, Appl1
142	32	72.7	243	2	US-08-916-902A-4	Sequence 4, Appl1	215	31	70.5	255	4	US-09-702-705-1806	Sequence 1806, Ap
143	32	72.7	243	2	US-09-213-389-4	Sequence 4, Appl1	216	31	70.5	255	4	US-09-736-457-1806	Sequence 1806, Ap
144	32	72.7	243	3	US-09-103-663-29	Sequence 29, Appl1	217	31	70.5	255	4	US-09-671-325-1806	Sequence 1806, Ap
145	32	72.7	243	3	US-09-103-663-33	Sequence 33, Appl1	218	31	70.5	261	4	US-09-205-258-505	Sequence 505, App
146	32	72.7	246	1	US-07-869-933-23	Sequence 23, Appl1	219	31	70.5	265	4	US-08-469-260A-77	Sequence 77, Appl1
147	32	72.7	246	3	US-09-103-663-23	Sequence 23, Appl1	220	31	70.5	265	4	US-08-488-446-77	Sequence 77, Appl1
148	32	72.7	306	2	US-08-560-098A-45	Sequence 45, Appl1	221	31	70.5	265	4	US-08-467-344A-77	Sequence 77, Appl1
149	32	72.7	331	2	US-08-560-098A-46	Sequence 46, Appl1	222	31	70.5	260	4	US-09-247-155-178	Sequence 178, App
150	32	72.7	427	4	US-09-506-066E-8	Sequence 8, Appl1	223	31	70.5	295	4	US-09-252-991A-19816	Sequence 19816, A
151	32	72.7	427	5	PCT-US95-04910-13	Sequence 13, Appl1	224	31	70.5	308	2	US-08-807-050-1	Sequence 1, Appl1
152	32	72.7	452	1	US-09-055-113-4	Sequence 4, Appl1	225	31	70.5	308	2	US-08-807-050-3	Sequence 3, Appl1
153	32	72.7	462	1	US-08-458-023B-2	Sequence 2, Appl1	226	31	70.5	308	2	US-08-807-050-4	Sequence 4, Appl1
154	32	72.7	463	3	US-09-111-556A-2	Sequence 2, Appl1	227	31	70.5	314	4	US-09-355-166-17	Sequence 17, Appl1
155	32	72.7	463	3	US-08-360-758-2	Sequence 2, Appl1	228	31	70.5	314	4	US-09-634-137-30	Sequence 30, Appl1
156	32	72.7	531	3	US-09-360-197-14	Sequence 14, Appl1	229	31	70.5	317	4	US-09-604-258-504	Sequence 504, App
157	32	72.7	571	4	US-09-252-991A-28549	Sequence 28549, A	230	31	70.5	319	4	US-09-252-991A-35635	Sequence 32635, A
158	32	72.7	622	4	US-09-338-352-4776	Sequence 4776, Ap	231	31	70.5	369	4	US-09-252-991A-20049	Sequence 20049, A
159	32	72.7	652	2	US-08-751-305-2	Sequence 2, Appl1	232	31	70.5	369	4	US-09-198-452A-521	Sequence 521, App
160	32	72.7	685	2	US-08-878-989-1	Sequence 1, Appl1	233	31	70.5	370	4	US-09-252-991A-27810	Sequence 27810, A
161	32	72.7	685	3	US-09-136-282-2	Sequence 2, Appl1	234	31	70.5	373	4	US-09-066-281B-19	Sequence 19, Appl1
162	32	72.7	685	3	US-09-272-796-1	Sequence 1, Appl1	235	31	70.5	373	4	US-09-468-433C-19	Sequence 19, Appl1
163	32	72.7	685	3	US-09-505-744-2	Sequence 2, Appl1	236	31	70.5	375	4	US-09-252-991A-24278	Sequence 24278, A
164	32	72.7	710	4	US-09-079-812E-2	Sequence 2, Appl1	237	31	70.5	397	4	US-09-252-991A-18219	Sequence 18219, A
165	32	72.7	723	4	US-09-434-408-2	Sequence 2, Appl1	238	31	70.5	398	1	US-08-471-791-20	Sequence 20, Appl1
166	32	72.7	760	4	US-09-252-991A-18711	Sequence 18711, A	239	31	70.5	398	4	US-08-926-522-6	Sequence 6, Appl1
167	32	72.7	774	4	US-09-252-991A-16789	Sequence 16789, A	240	31	70.5	398	5	PCT-US91-01746-20	Sequence 20, Appl1
168	32	72.7	965	4	US-09-252-991A-24038	Sequence 24038, A	241	31	70.5	445	4	US-09-252-991A-17013	Sequence 17013, A
169	32	72.7	972	4	US-09-710-092-2	Sequence 2, Appl1	242	31	70.5	446	1	US-07-952-800-4	Sequence 4, Appl1
170	32	72.7	1056	4	US-09-710-092-6	Sequence 6, Appl1	243	31	70.5	448	4	US-08-216-592A-2	Sequence 2, Appl1
171	32	72.7	1130	4	US-09-976-594-880	Sequence 280, App	244	31	70.5	463	4	US-09-540-236-2942	Sequence 2942, Ap
172	32	72.7	1270	4	US-09-710-092-10	Sequence 10, Appl1	245	31	70.5	466	4	US-09-604-107A-8	Sequence 8, Appl1
173	32	72.7	1294	2	US-08-819-288-3	Sequence 3, Appl1	246	31	70.5	501	3	US-08-857-076-102	Sequence 102, App

247	31	70.5	510	5	PCT-US96-03916-3	Sequence 3, Appl1	320	30	68.2	217	4	US-08-467-602-399	Sequence 399, App
248	31	70.5	559	4	US-09-364-206-47	Sequence 47, Appl	321	30	68.2	217	4	US-08-467-602-413	Sequence 413, App
249	31	70.5	560	2	US-08-808-931-18	Sequence 18, Appl	322	30	68.2	224	3	US-08-974-022-50	Sequence 50, Appl
250	31	70.5	560	3	US-08-808-323-18	Sequence 18, Appl	323	30	68.2	224	3	US-08-795-445A-50	Sequence 50, Appl
251	31	70.5	560	3	US-09-050-603A-18	Sequence 18, Appl	324	30	68.2	224	3	US-08-795-447A-50	Sequence 50, Appl
252	31	70.5	560	3	US-09-102-420B-18	Sequence 18, Appl	325	30	68.2	224	3	US-08-974-186-50	Sequence 50, Appl
253	31	70.5	560	4	US-09-497-698-18	Sequence 18, Appl	326	30	68.2	224	3	US-08-795-446B-50	Sequence 50, Appl
254	31	70.5	575	4	US-09-171-461-30	Sequence 30, Appl	327	30	68.2	224	4	US-08-706-945B-137	Sequence 137, App
255	31	70.5	593	4	US-09-252-991A-23251	Sequence 23251, A	328	30	68.2	224	4	US-08-577-788B-51	Sequence 51, Appl
256	31	70.5	623	5	PCT-US96-03916-64	Sequence 64, Appl1	329	30	68.2	234	4	US-09-489-039A-12555	Sequence 12555, A
257	31	70.5	731	3	US-09-115-446-2	Sequence 2, Appl1	330	30	68.2	239	3	US-08-470-335-219	Sequence 219, App
258	31	70.5	731	4	US-09-275-252A-10	Sequence 10, Appl1	331	30	68.2	239	3	US-08-470-339-219	Sequence 219, App
259	31	70.5	731	4	US-09-565-590-2	Sequence 2, Appl1	332	30	68.2	250	4	US-09-252-991A-18515	Sequence 18515, A
260	31	70.5	755	4	US-09-252-991A-28386	Sequence 28386, A	333	30	68.2	253	1	US-08-399-696-4	Sequence 4, Appl1
261	31	70.5	796	4	US-09-540-236-22655	Sequence 22655, Ap	334	30	68.2	253	1	US-08-399-696-118	Sequence 118, App
262	31	70.5	865	4	US-09-281-766-19	Sequence 19, Appl	335	30	68.2	258	4	US-09-252-991A-20810	Sequence 20810, A
263	31	70.5	933	2	US-08-682-847-2	Sequence 2, Appl1	336	30	68.2	279	4	US-09-399-588-4	Sequence 4, Appl1
264	31	70.5	1377	2	US-08-308-818-4	Sequence 4, Appl1	337	30	68.2	281	2	US-08-810-453-2	Sequence 2, Appl1
265	31	70.5	1440	3	US-09-357-251-37	Sequence 37, Appl1	338	30	68.2	281	3	US-08-815-190A-2	Sequence 2, Appl1
266	31	70.5	1464	3	US-09-445-360-2	Sequence 2, Appl1	339	30	68.2	281	3	US-09-490-640-25	Sequence 25, Appl
267	31	70.5	1464	4	US-09-713-273A-21	Sequence 21, Appl	340	30	68.2	281	3	US-09-479-524-3	Sequence 3, Appl1
268	31	70.5	1469	4	US-09-713-273A-16	Sequence 16, Appl	341	30	68.2	281	4	US-08-339-214-8	Sequence 8, Appl1
269	31	70.5	1512	4	US-09-443-184-48	Sequence 48, Appl	342	30	68.2	281	4	US-08-339-214-30	Sequence 30, Appl
270	31	70.5	2584	3	US-08-336-135-4	Sequence 4, Appl1	343	30	68.2	281	4	US-09-157-864-10	Sequence 6, Appl1
271	31	70.5	2588	3	US-08-336-135-2	Sequence 2, Appl1	344	30	68.2	281	4	US-09-157-864-10	Sequence 10, Appl
272	30	68.2	11	1	US-08-330-047-22	Sequence 22, Appl	345	30	68.2	281	4	US-09-006-755B-1	Sequence 1, Appl1
273	30	68.2	11	2	US-08-747-137-102	Sequence 102, App	346	30	68.2	281	4	US-09-027-287-6	Sequence 6, Appl1
274	30	68.2	11	4	US-09-140-084-6	Sequence 6, Appl1	347	30	68.2	281	4	US-09-252-656B-6	Sequence 6, Appl1
275	30	68.2	11	4	US-09-724-297-6	Sequence 6, Appl1	348	30	68.2	281	4	US-08-968-686-12	Sequence 12, Appl
276	30	68.2	11	6	5204096-29	Patent No. 5204096	349	30	68.2	281	4	US-09-588-947A-6	Sequence 6, Appl1
277	30	68.2	15	4	US-08-339-211A-64	Sequence 64, Appl1	350	30	68.2	281	4	US-09-589-286A-6	Sequence 6, Appl1
278	30	68.2	20	3	US-08-649-100-1	Sequence 1, Appl1	351	30	68.2	281	4	US-09-523-323-6	Sequence 6, Appl1
279	30	68.2	24	1	US-08-399-696-106	Sequence 106, App	352	30	68.2	281	5	PCT-US95-00362-2	Sequence 25, Appl
280	30	68.2	24	1	US-08-399-696-111	Sequence 111, App	353	30	68.2	281	5	US-08-815-190A-16	Sequence 16, Appl
281	30	68.2	32	4	US-09-006-755B-5	Sequence 5, Appl1	354	30	68.2	287	3	US-08-467-963C-8	Sequence 8, Appl1
282	30	68.2	36	4	US-09-006-755B-4	Sequence 4, Appl1	355	30	68.2	298	2	US-08-467-963C-8	Sequence 8, Appl1
283	30	68.2	43	4	US-09-122-144-7	Sequence 7, Appl1	356	30	68.2	298	2	US-08-338-189B-8	Sequence 8, Appl1
284	30	68.2	55	4	US-09-314-268-147	Sequence 147, App	357	30	68.2	298	3	US-08-852-344D-8	Sequence 8, Appl1
285	30	68.2	92	2	US-08-459-568-81	Sequence 81, Appl	358	30	68.2	298	3	US-08-344-639B-8	Sequence 8, Appl1
286	30	68.2	92	2	US-08-399-411-81	Sequence 81, Appl	359	30	68.2	298	3	US-08-467-969A-8	Sequence 8, Appl1
287	30	68.2	92	3	US-08-516-859A-81	Sequence 81, Appl	360	30	68.2	298	3	US-08-467-961A-8	Sequence 8, Appl1
288	30	68.2	92	4	US-09-586-472-81	Sequence 81, Appl	361	30	68.2	298	3	US-08-001-554A-8	Sequence 8, Appl1
289	30	68.2	92	4	US-09-528-706-81	Sequence 81, Appl	362	30	68.2	307	4	US-09-252-991A-29783	Sequence 29783, A
290	30	68.2	92	4	US-09-621-976-4360	Sequence 4360, Ap	363	30	68.2	322	4	US-09-252-991A-31782	Sequence 31782, A
291	30	68.2	97	4	US-09-252-991A-25508	Sequence 25508, A	364	30	68.2	333	4	US-09-489-039A-7941	Sequence 7941, Ap
292	30	68.2	119	3	US-09-012-515A-24	Sequence 24, Appl	365	30	68.2	346	4	US-09-455-921A-2	Sequence 2, Appl1
293	30	68.2	119	3	US-08-360-144A-24	Sequence 24, Appl	366	30	68.2	346	4	US-09-468-433C-22	Sequence 22, Appl
294	30	68.2	119	4	US-09-012-504A-24	Sequence 24, Appl	367	30	68.2	348	1	US-07-989-845-2	Sequence 2, Appl1
295	30	68.2	119	4	US-09-012-399A-24	Sequence 2, Appl1	368	30	68.2	348	5	PCT-US93-11298-2	Sequence 2, Appl1
296	30	68.2	142	3	US-08-350-468-2	Sequence 2, Appl1	369	30	68.2	359	4	US-09-328-352-8005	Sequence 8005, Ap
297	30	68.2	149	3	US-08-584-031-17	Sequence 17, Appl	370	30	68.2	376	3	US-08-751-512-8	Sequence 8, Appl1
298	30	68.2	152	4	US-09-596-118B-8	Sequence 8, Appl1	371	30	68.2	386	4	US-09-252-991A-24221	Sequence 24221, A
299	30	68.2	154	3	US-08-716-190-6	Sequence 6, Appl1	372	30	68.2	391	1	US-07-995-657-6	Sequence 6, Appl1
300	30	68.2	157	4	US-09-252-991A-30836	Sequence 30836, A	373	30	68.2	391	1	US-08-474-587-6	Sequence 6, Appl1
301	30	68.2	158	1	US-08-706-214-3	Sequence 3, Appl1	374	30	68.2	398	4	US-09-252-991A-20748	Sequence 20748, A
302	30	68.2	158	1	US-08-318-947A-5	Sequence 2, Appl1	375	30	68.2	406	4	US-09-252-991A-32895	Sequence 32895, A
303	30	68.2	158	1	US-08-318-947A-5	Sequence 2, Appl1	376	30	68.2	419	3	US-08-509-024-7	Sequence 7, Appl1
304	30	68.2	158	2	US-08-795-503-2	Sequence 5, Appl1	377	30	68.2	419	4	US-09-333-279-7	Sequence 7, Appl1
305	30	68.2	158	2	US-08-795-503-5	Sequence 5, Appl1	378	30	68.2	419	4	US-09-631-780-7	Sequence 7, Appl1
306	30	68.2	158	2	US-08-486-663A-13	Sequence 13, Appl	379	30	68.2	427	3	US-09-086-483A-4	Sequence 4, Appl1
307	30	68.2	158	3	US-08-767-942A-13	Sequence 13, Appl	380	30	68.2	427	3	US-09-041-886-2	Sequence 2, Appl1
308	30	68.2	158	3	US-08-350-468-4	Sequence 4, Appl1	381	30	68.2	427	3	US-09-006-351A-5	Sequence 5, Appl1
309	30	68.2	158	4	US-09-886-319A-69	Sequence 69, Appl1	382	30	68.2	427	4	US-08-467-602-189	Sequence 189, App
310	30	68.2	158	4	US-09-886-319A-70	Sequence 70, Appl	383	30	68.2	427	4	US-09-573-986-5	Sequence 5, Appl1
311	30	68.2	178	5	PCT-US95-06722-24	Sequence 24, Appl	384	30	68.2	427	4	US-09-580-212-4	Sequence 4, Appl1
312	30	68.2	178	5	US-08-339-214-84	Sequence 84, Appl	385	30	68.2	427	4	US-09-769-402-4	Sequence 4, Appl1
313	30	68.2	179	3	US-08-649-100-9	Sequence 9, Appl1	386	30	68.2	427	4	US-09-748-537-13	Sequence 13, Appl
314	30	68.2	179	4	US-08-339-214-6	Sequence 6, Appl1	387	30	68.2	429	4	US-09-252-991A-25920	Sequence 25920, A
315	30	68.2	196	4	US-09-252-991A-26335	Sequence 26335, A	388	30	68.2	430	4	US-08-467-602-187	Sequence 187, App
316	30	68.2	217	3	US-08-341-018-48	Sequence 48, Appl	389	30	68.2	436	4	US-08-467-602-201	Sequence 201, App
317	30	68.2	217	3	US-08-341-018-50	Sequence 50, Appl	390	30	68.2	439	4	US-08-467-602-198	Sequence 198, App
318	30	68.2	217	3	US-08-470-335-205	Sequence 205, App	391	30	68.2	450	4	US-08-467-602-195	Sequence 195, App
319	30	68.2	217	3	US-08-470-339-205	Sequence 205, App	392	30	68.2	452	4	US-09-198-452A-808	Sequence 808, App

393	30	68.2	455	4	US-09-527-236A-4	Sequence 4, Appl1	466	30	68.2	650	4	US-08-467-602-280	Sequence 280, App
394	30	68.2	455	4	US-09-756-854-4	Sequence 4, Appl1	467	30	68.2	653	4	US-08-467-602-212	Sequence 212, App
395	30	68.2	456	4	US-09-3328-352-6946	Sequence 6946, App	468	30	68.2	653	4	US-08-467-602-275	Sequence 275, App
396	30	68.2	459	4	US-08-467-602-204	Sequence 204, App	469	30	68.2	659	4	US-08-467-602-229	Sequence 229, App
397	30	68.2	474	4	US-08-467-602-190	Sequence 190, App	470	30	68.2	659	4	US-08-467-602-292	Sequence 292, App
398	30	68.2	477	4	US-08-467-602-185	Sequence 185, App	471	30	68.2	660	4	US-08-467-602-264	Sequence 264, App
399	30	68.2	483	4	US-08-467-602-202	Sequence 202, App	472	30	68.2	660	4	US-09-252-991A-22842	Sequence 22842, A
400	30	68.2	486	4	US-08-467-602-199	Sequence 199, App	473	30	68.2	662	4	US-08-467-602-226	Sequence 226, App
401	30	68.2	497	4	US-08-467-602-193	Sequence 193, App	474	30	68.2	662	4	US-08-467-602-289	Sequence 289, App
402	30	68.2	498	3	US-09-238-303-8	Sequence 8, Appl1	475	30	68.2	669	1	US-07-847-743B-8	Sequence 8, Appl1
403	30	68.2	498	4	US-09-946-239-8	Sequence 8, Appl1	476	30	68.2	669	1	US-08-456-201-8	Sequence 13, Appl1
404	30	68.2	506	4	US-08-467-602-205	Sequence 205, App	477	30	68.2	669	1	US-08-456-201-13	Sequence 8, Appl1
405	30	68.2	528	3	US-08-928-213B-8	Sequence 8, Appl1	478	30	68.2	669	1	US-08-330-161-11	Sequence 13, Appl1
406	30	68.2	560	3	US-08-341-018-58	Sequence 58, Appl1	479	30	68.2	669	2	US-08-456-241-8	Sequence 11, Appl1
407	30	68.2	560	3	US-08-470-335-194	Sequence 194, App	480	30	68.2	669	2	US-08-456-241-13	Sequence 8, Appl1
408	30	68.2	560	3	US-08-470-339-194	Sequence 194, App	481	30	68.2	669	2	US-08-456-241-13	Sequence 13, Appl1
409	30	68.2	560	4	US-08-467-602-388	Sequence 388, App	482	30	68.2	669	2	US-08-440-401-11	Sequence 11, Appl1
410	30	68.2	569	4	US-08-467-602-377	Sequence 237, App	483	30	68.2	669	2	US-08-419-878A-11	Sequence 11, Appl1
411	30	68.2	572	4	US-08-467-602-335	Sequence 235, App	484	30	68.2	669	4	US-09-173-480-11	Sequence 11, Appl1
412	30	68.2	578	4	US-08-467-602-249	Sequence 249, App	485	30	68.2	669	4	US-08-467-602-273	Sequence 273, App
413	30	68.2	581	4	US-08-467-602-246	Sequence 246, App	486	30	68.2	669	5	PCT-US92-04295A-8	Sequence 8, Appl1
414	30	68.2	582	4	US-08-467-602-243	Sequence 243, App	487	30	68.2	669	5	PCT-US92-04295A-13	Sequence 13, Appl1
415	30	68.2	593	4	US-08-637-670-38	Sequence 38, Appl1	488	30	68.2	673	4	US-08-467-602-220	Sequence 220, App
416	30	68.2	601	4	US-08-467-602-252	Sequence 252, App	489	30	68.2	673	4	US-08-467-602-283	Sequence 283, App
417	30	68.2	602	1	US-08-168-091A-2	Sequence 2, Appl1	490	30	68.2	682	4	US-08-467-602-232	Sequence 232, App
418	30	68.2	602	1	US-08-428-926-5	Sequence 5, Appl1	491	30	68.2	682	4	US-08-467-602-295	Sequence 295, App
419	30	68.2	602	1	US-08-428-927-5	Sequence 5, Appl1	492	30	68.2	682	4	US-08-467-602-229	Sequence 259, App
420	30	68.2	602	1	US-08-428-398-5	Sequence 5, Appl1	493	30	68.2	687	4	US-08-467-602-254	Sequence 254, App
421	30	68.2	602	1	US-08-339-517-5	Sequence 5, Appl1	494	30	68.2	687	4	US-08-467-602-271	Sequence 271, App
422	30	68.2	603	4	US-08-467-602-216	Sequence 216, App	495	30	68.2	685	4	US-09-684-708A-5	Sequence 5, Appl1
423	30	68.2	603	4	US-08-467-602-219	Sequence 219, App	496	30	68.2	686	4	US-08-467-602-268	Sequence 268, App
424	30	68.2	606	4	US-08-467-602-214	Sequence 214, App	497	30	68.2	700	4	US-09-684-708A-2	Sequence 2, Appl1
425	30	68.2	606	4	US-08-467-602-277	Sequence 277, App	498	30	68.2	707	4	US-08-467-602-262	Sequence 262, App
426	30	68.2	612	4	US-08-467-602-228	Sequence 228, App	499	30	68.2	716	4	US-08-467-602-274	Sequence 274, App
427	30	68.2	612	4	US-08-467-602-291	Sequence 291, App	500	30	68.2	725	4	US-09-252-991A-23752	Sequence 23752, A
428	30	68.2	615	4	US-08-467-602-225	Sequence 225, App							
429	30	68.2	615	4	US-08-467-602-288	Sequence 288, App							
430	30	68.2	616	4	US-08-467-602-238	Sequence 238, App							
431	30	68.2	619	4	US-08-467-602-233	Sequence 233, App							
432	30	68.2	625	1	US-07-847-743B-26	Sequence 26, Appl1							
433	30	68.2	625	1	US-08-456-201-26	Sequence 26, Appl1							
434	30	68.2	625	2	US-08-456-241-26	Sequence 26, Appl1							
435	30	68.2	625	4	US-08-467-602-250	Sequence 250, App							
436	30	68.2	625	5	PCT-US92-04295A-26	Sequence 25, Appl1							
437	30	68.2	626	4	US-08-467-602-222	Sequence 222, App							
438	30	68.2	626	4	US-08-467-602-285	Sequence 285, App							
439	30	68.2	628	4	US-08-467-602-247	Sequence 247, App							
440	30	68.2	635	4	US-08-467-602-231	Sequence 231, App							
441	30	68.2	635	4	US-08-467-602-294	Sequence 294, App							
442	30	68.2	637	1	US-07-847-743B-28	Sequence 28, Appl1							
443	30	68.2	637	1	US-08-456-201-28	Sequence 28, Appl1							
444	30	68.2	637	2	US-08-456-241-28	Sequence 28, Appl1							
445	30	68.2	637	4	US-08-467-602-258	Sequence 258, App							
446	30	68.2	637	5	PCT-US92-04295A-28	Sequence 28, Appl1							
447	30	68.2	639	4	US-08-467-602-241	Sequence 241, App							
448	30	68.2	640	4	US-08-467-602-256	Sequence 256, App							
449	30	68.2	641	4	US-09-489-039A-8696	Sequence 8696, App							
450	30	68.2	645	1	US-07-847-743B-27	Sequence 27, Appl1							
451	30	68.2	645	1	US-08-456-201-27	Sequence 27, Appl1							
452	30	68.2	645	1	US-08-428-926-4	Sequence 4, Appl1							
453	30	68.2	645	1	US-08-428-927-4	Sequence 4, Appl1							
454	30	68.2	645	1	US-08-428-928-4	Sequence 4, Appl1							
455	30	68.2	645	1	US-08-339-517-4	Sequence 4, Appl1							
456	30	68.2	645	2	US-08-456-241-27	Sequence 27, Appl1							
457	30	68.2	645	3	US-08-753-007A-10	Sequence 10, Appl1							
458	30	68.2	645	3	US-09-398-496-10	Sequence 10, Appl1							
459	30	68.2	645	3	US-09-020-880-93	Sequence 93, Appl1							
460	30	68.2	645	3	US-09-101-544-93	Sequence 93, Appl1							
461	30	68.2	645	5	PCT-US92-04295A-27	Sequence 27, Appl1							
462	30	68.2	646	5	US-08-467-602-270	Sequence 270, App							
463	30	68.2	649	4	US-08-467-602-253	Sequence 253, App							
464	30	68.2	649	4	US-08-467-602-267	Sequence 267, App							
465	30	68.2	650	4	US-08-467-602-217	Sequence 217, App							

466	30	68.2	650	4	US-08-467-602-280	Sequence 280, App
467	30	68.2	653	4	US-08-467-602-212	Sequence 212, App
468	30	68.2	653	4	US-08-467-602-275	Sequence 275, App
469	30	68.2	659	4	US-08-467-602-229	Sequence 229, App
470	30	68.2	659	4	US-08-467-602-292	Sequence 292, App
471	30	68.2	660	4	US-08-467-602-264	Sequence 264, App
472	30	68.2	660	4	US-09-252-991A-22842	Sequence 22842, A
473	30	68.2	662	4	US-08-467-602-226	Sequence 226, App
474	30	68.2	662	4	US-08-467-602-289	Sequence 289, App
475	30	68.2	669	1	US-07-847-743B-8	Sequence 8, Appl1
476	30	68.2	669	1	US-08-456-201-8	Sequence 13, Appl1
477	30	68.2	669	1	US-08-456-201-13	Sequence 8, Appl1
478	30	68.2	669	1	US-08-330-161-11	Sequence 13, Appl1
479	30	68.2	669	2	US-08-456-241-8	Sequence 11, Appl1
480	30	68.2	669	2	US-08-456-241-13	Sequence 8, Appl1
481	30	68.2	669	2	US-08-456-241-13	Sequence 13, Appl1
482	30	68.2	669	2	US-08-440-401-11	Sequence 11, Appl1
483	30	68.2	669	2	US-08-419-878A-11	Sequence 11, Appl1
484	30	68.2	669	4	US-09-173-480-11	Sequence 11, Appl1
485	30	68.2	669	4	US-08-467-602-273	Sequence 273, App
486	30	68.2	669	5	PCT-US92-04295A-8	Sequence 8, Appl1
487	30	68.2	669	5	PCT-US92-04295A-13	Sequence 13, Appl1
488	30	68.2	673	4	US-08-467-602-220	Sequence 220, App
489	30	68.2	673	4	US-08-467-602-283	Sequence 283, App
490	30	68.2	682	4	US-08-467-602-232	Sequence 232, App
491	30	68.2	682	4	US-08-467-602-295	Sequence 295, App
492	30	68.2	682	4	US-08-467-602-229	Sequence 259, App
493	30	68.2	687	4	US-08-467-602-254	Sequence 254, App
494	30	68.2	687	4	US-08-467-602-271	Sequence 271, App
495	30	68.2	685	4	US-09-684-708A-5	Sequence 5, Appl1
496	30	68.2	686	4	US-08-467-602-268	Sequence 268, App
497	30	68.2	700	4	US-09-684-708A-2	Sequence 2, Appl1
498	30	68.2	707	4	US-08-467-602-262	Sequence 262, App
499	30	68.2	716	4	US-08-467-602-274	Sequence 274, App
500	30	68.2	725	4	US-09-252-991A-23752	Sequence 23752, A

## ALIGNMENTS

```

RESULT 1
US-09-101-272G-77
; Sequence 77, Application US/09101272G
; Patent No. 6509445
; GENERAL INFORMATION:
; APPLICANT: Nisissin Food Products Co., Ltd.
; TITLE OF INVENTION: CANCEROUS METASTASIS INHIBITOR
; FILE REFERENCE: Q50979
; CURRENT APPLICATION NUMBER: US/09/101, 272G
; CURRENT FILING DATE: 1998-07-08
; PRIOR APPLICATION NUMBER: JP 1059/1996
; PRIOR FILING DATE: 1996-01-08
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: residues 135-143 of the ATF domain of uPA
US-09-101-272G-77

Query Match      100.0%; Score 44; DB 4; Length 9;
Best Local Similarity 100.0%; Pred No. 3e+05;
Matches      8; Conservative 0; Mismatches 0; Gaps 0;
Qy      1 KPSPPEE 8
      |||||
Db      2 KPSPPEE 9

RESULT 2

```



US-07-603-675-1  
; Sequence 1, Application US/07603675  
; Patent No. 5416006  
; GENERAL INFORMATION:  
; APPLICANT: Biasi, Francesco  
; APPLICANT: Stoppelli, Maria P  
; APPLICANT: Mastrolicola, Maria R  
; APPLICANT: Weindler, Karen G  
; APPLICANT: Correas, Isabel  
; TITLE OF INVENTION: MODIFICATION OF PLASMINOGEN ACTIVATORS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: COOPER & DUNHAM  
; STREET: 30 ROCKEFELLER PLAZA  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: U.S.A.  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/603,675  
; FILING DATE: 19911218  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/DK90/00096  
; FILING DATE: 11-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P  
; REGISTRATION NUMBER: 28,678  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 977-9550  
; TELEFAX: (212) 644-0525  
; TELEX: (212) 422523 COOP UI  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: N  
; FRAGMENT TYPE: Internal  
; US-07-603-675-1  
Query Match 100.0%; Score 44; DB 1; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KPSSPPEE 8  
Db 1 KPSSPPEE 8  
RESULT 3  
US-07-603-675-8  
; Sequence 8, Application US/07603675  
; Patent No. 5416006  
; GENERAL INFORMATION:  
; APPLICANT: Biasi, Francesco  
; APPLICANT: Stoppelli, Maria P  
; APPLICANT: Mastrolicola, Maria R  
; APPLICANT: Weindler, Karen G  
; APPLICANT: Correas, Isabel  
; TITLE OF INVENTION: MODIFICATION OF PLASMINOGEN ACTIVATORS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: COOPER & DUNHAM  
; STREET: 30 ROCKEFELLER PLAZA  
; CITY: NEW YORK

STATE: NEW YORK  
; COUNTRY: U.S.A.  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/603,675  
; FILING DATE: 19911218  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/DK90/00096  
; FILING DATE: 11-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P  
; REGISTRATION NUMBER: 28,678  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 977-9550  
; TELEFAX: (212) 644-0525  
; TELEX: (212) 422523 COOP UI  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 11 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: N  
; FRAGMENT TYPE: Internal  
; US-07-603-675-8  
Query Match 100.0%; Score 44; DB 1; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.66;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KPSSPPEE 8  
Db 4 KPSSPPEE 11

RESULT 4  
US-09-101-272G-38  
; Sequence 38, Application US/09101272G  
; Patent No. 6509445  
; GENERAL INFORMATION:  
; APPLICANT: Nissin Food Products Co., Ltd.  
; TITLE OF INVENTION: CANCEROUS METASTASIS INHIBITOR  
; FILE REFERENCE: 050979  
; CURRENT APPLICATION NUMBER: US/09/101,272G  
; CURRENT FILING DATE: 1998-07-08  
; PRIOR APPLICATION NUMBER: JP 1059/1996  
; PRIOR FILING DATE: 1996-01-08  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 38  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Intervening sequence between formula 1 and formula 2  
; US-09-101-272G-38  
Query Match 100.0%; Score 44; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KPSSPPEE 8  
Db 5 KPSSPPEE 12

RESULT 5  
US-08-142-590B-25  
; Sequence 25, Application US/08142590B  
; Patent No. 6120765  
; GENERAL INFORMATION:  
; APPLICANT: HIBINO, Tashihiko; TAKAHASHI, Tadahiko; HORII, Izumi; and GOETINCK,  
; TITLE OF INVENTION: UROKINASE PLASMINOGEN ACTIVATOR FRAGMENTS  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD, LLP  
; STREET: 28 State Street  
; City: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/142,590B  
; FILING DATE: 25-OCT-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/042,318  
; FILING DATE: 02-APR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Myers, Paul L.  
; REGISTRATION NUMBER: 35,965  
; REFERENCE/DOCKET NUMBER: MGP-009CP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 227-5941  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 157 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: internal  
; US-08-142-590B-25

Query Match 100.0%; Score 44; DB 3; Length 157;  
Best Local Similarity 100.0%; Pred. No. 8.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
Db 136 KPSSPPE 143

RESULT 6  
US-09-101-272G-73  
; Sequence 73, Application US/09101272G  
; Patent No. 6509445  
; GENERAL INFORMATION:  
; APPLICANT: Nissin Food Products Co., Ltd.  
; TITLE OF INVENTION: CANCEROUS METASTASIS INHIBITOR  
; FILE REFERENCE: 050979  
; CURRENT APPLICATION NUMBER: US/09/101,272G  
; CURRENT FILING DATE: 1998-07-08  
; PRIOR APPLICATION NUMBER: JP 1059/1996  
; PRIOR FILING DATE: 1996-01-08  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 73  
; LENGTH: 200  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: ATF domain of uPA  
; US-09-101-272G-73

Query Match 100.0%; Score 44; DB 4; Length 200;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
Db 156 KPSSPPE 163

RESULT 7  
US-09-101-272G-98  
; Sequence 98, Application US/09101272G  
; Patent No. 6509445  
; GENERAL INFORMATION:  
; APPLICANT: Nissin Food Products Co., Ltd.  
; TITLE OF INVENTION: CANCEROUS METASTASIS INHIBITOR  
; FILE REFERENCE: 050979  
; CURRENT APPLICATION NUMBER: US/09/101,272G  
; CURRENT FILING DATE: 1998-07-08  
; PRIOR APPLICATION NUMBER: JP 1059/1996  
; PRIOR FILING DATE: 1996-01-08  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 98  
; LENGTH: 208  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: ATFHI-ML chimeric protein  
; US-09-101-272G-98

Query Match 100.0%; Score 44; DB 4; Length 208;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
Db 137 KPSSPPE 144

RESULT 8  
US-08-093-741-83  
; Sequence 83, Application US/08093741  
; Patent No. 5681721  
; GENERAL INFORMATION:  
; APPLICANT: STEFFENS, GERD J.  
; APPLICANT: WENNDT, STEPHAN  
; APPLICANT: SCHNEIDER, JOHANNES  
; APPLICANT: HEINZEL-WIELAND, REGINA  
; TITLE OF INVENTION: BIFUNCTIONAL UROKINASE VARIANTS WITH  
; TITLE OF INVENTION: IMPROVED FIBRINOLYTIC CHARACTERISTICS AND THROMBIN  
; TITLE OF INVENTION: INHIBITING EFFECT  
; NUMBER OF SEQUENCES: 83  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Evenson, McKeown, Edwards & Lenahan  
; STREET: 1200 G Street, N. W. Suite 700  
; City: Washington, D.C.  
; COUNTRY: U.S.  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Releasee #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/093,741  
; FILING DATE: 20-JUL-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P43 23 754.1  
; FILING DATE: 15-JUL-1993

ATTORNEY/AGENT INFORMATION:  
NAME: EVANS, Joseph D.  
REGISTRATION NUMBER: 26,269  
REFERENCE/DOCKET NUMBER: 148/41345  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-8800  
TELEFAX: (202) 628-8844  
INFORMATION FOR SEQ ID NO: 83:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 365 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-093-741-83

Query Match 100.0%; Score 44; DB 1; Length 365;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8  
Db 90 KPSSPPEE 97

RESULT 9  
US-08-720-012-83  
Sequence 83, Application US/08720012  
Patent No. 5747291  
GENERAL INFORMATION:  
APPLICANT: STEFFENS, Gerd J.  
APPLICANT: WENNDT, STEPHAN  
APPLICANT: SCHNEIDER, JOHANNES  
APPLICANT: HEINZEL-WIELAND, REGINA  
APPLICANT: SAUNDERS, DEREK J.  
TITLE OF INVENTION: BIFUNCTIONAL UROKINASE VARIANTS WITH  
TITLE OF INVENTION: IMPROVED FIBRINOLYTIC CHARACTERISTICS AND THROMBIN  
NUMBER OF SEQUENCES: 83  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Evenson, McKeown, Edwards & Lenahan  
STREET: 1200 G Street, N. W. Suite 700  
CITY: Washington, D.C.  
COUNTRY: U.S.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/720,012  
FILING DATE: 27-SEP-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/093,741  
FILING DATE: 20-JUL-1993  
APPLICATION NUMBER: DE P43 23 754.1  
FILING DATE: 15-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: EVANS, Joseph D.  
REGISTRATION NUMBER: 26,269  
REFERENCE/DOCKET NUMBER: 148/41345  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-8800  
TELEFAX: (202) 628-8844  
INFORMATION FOR SEQ ID NO: 83:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 365 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-720-012-83

Query Match 100.0%; Score 44; DB 1; Length 365;  
Best Local Similarity 100.0%; Pred. No. 20;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 KPSSPPEE 8  
Db 90 KPSSPPEE 97

RESULT 10  
US-08-560-098A-44  
Sequence 44, Application US/08560098A  
Patent No. 5976841  
GENERAL INFORMATION:  
APPLICANT: WENNDT, Stephan  
APPLICANT: HEINZEL-WIELAND, Regina  
APPLICANT: STEFFENS, Gerd Josef  
TITLE OF INVENTION: Proteins having Fibrinolytic and  
TITLE OF INVENTION: Coagulation-inhibiting Properties  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Evenson, McKeown, Edwards & Lenahan  
STREET: 1200 G Street, N.W., Suite 700  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/560,098A  
FILING DATE: 17-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: P 44 40 892.7  
FILING DATE: 17-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: EVANS, Joseph D.  
REGISTRATION NUMBER: 26,269  
REFERENCE/DOCKET NUMBER: 148/42448  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-8800  
TELEFAX: (202) 628-8844  
INFORMATION FOR SEQ ID NO: 44:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 393 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-560-098A-44

Query Match 100.0%; Score 44; DB 2; Length 393;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8  
Db 91 KPSSPPEE 98

RESULT 11  
US-08-967-024C-24  
Sequence 24, Application US/08967024C  
Patent No. 6133011  
GENERAL INFORMATION:  
APPLICANT: STEFFENS, Gerd Josef  
APPLICANT: JANOSHA, Biko  
APPLICANT: HEINZEL-WIELAND, Regina  
TITLE OF INVENTION: Chimeric Proteins having Fibrinolytic  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Evenson, McKeown, Edwards & Lenahan  
STREET: 1200 G Street, N.W., Suite 700  
CITY: Washington  
STATE: DC  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/967,024C  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: P 44 42 665.8  
FILING DATE: 30-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: EVANS, Joseph D.  
REGISTRATION NUMBER: 26,269  
REFERENCE/DOCKET NUMBER: 148/42444  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-8800  
TELEFAX: (202) 628-8844  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 393 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-967-024C-24

Query Match 100.0%; Score 44; DB 3; Length 393;  
Best local Similarity 100.0%; Pred. No. 22;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8  
|||||  
Db 91 KPSSPPEE 98

RESULT 12  
US-08-967-024C-25  
Sequence 25, Application US/08967024C  
Patent No. 6133011  
GENERAL INFORMATION:  
APPLICANT: MNENDT, Stephan  
APPLICANT: STEFFENS, Geird Josef  
APPLICANT: JANCOCHA, Elke  
APPLICANT: HEINZEL-WIELAND, Regina  
TITLE OF INVENTION: Chimeric Proteins having fibrinolytic  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Evenson, McKeown, Edwards & Lenahan  
STREET: 1200 G Street, N.W., Suite 700  
CITY: Washington  
STATE: DC  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/967,024C  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: P 44 42 665.8  
FILING DATE: 30-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: EVANS, Joseph D.

REGISTRATION NUMBER: 26,269  
REFERENCE/DOCKET NUMBER: 148/42444  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-8800  
TELEFAX: (202) 628-8844  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 393 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-967-024C-25

Query Match 100.0%; Score 44; DB 3; Length 393;  
Best local Similarity 100.0%; Pred. No. 22;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8  
|||||  
Db 91 KPSSPPEE 98

RESULT 13  
US-08-087-163-1  
Sequence 1, Application US/08087163  
Patent No. 5472692  
GENERAL INFORMATION:  
APPLICANT: Liu, Jian-Ning  
APPLICANT: Gurewicz, Victor  
TITLE OF INVENTION: PRO-UKINKINASE MUTANTS  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 50Z or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/087,163  
FILING DATE: 07/02/93  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Faasse, J. Peter  
REGISTRATION NUMBER: 32,983  
REFERENCE/DOCKET NUMBER: 04353/003001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 411  
TYPE: amino acid  
STRANDEDNESS: N/A  
TOPOLOGY: N/A  
US-08-087-163-1

Query Match 100.0%; Score 44; DB 1; Length 411;  
Best local Similarity 100.0%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8  
|||||

Db 136 KPSSPPE 143

## RESULT 14

US-08-286-748B-18  
; Sequence 18, Application US/08286748B

; Patent No. 5759542

; GENERAL INFORMATION:

; APPLICANT: Victor Gurewlich

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DELIVERY

; TITLE OF INVENTION: OF DRUGS BY PLATELETS FOR THE TREATMENT OF

; TITLE OF INVENTION: CARDIOVASCULAR AND OTHER DISEASES

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish &amp; Richardson

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; COMPUTER: IBM PS/2 Model 50Z or 55SX

; OPERATING SYSTEM: MS-DOS (Version 5.0)

; SOFTWARE: WordPerfect (Version 5.1)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/286,748B

; FILING DATE: August 5, 1994

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: J. Peter Paease

; REGISTRATION NUMBER: 32,983

; REFERENCE/DOCKET NUMBER: 04547/013001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 542-5070

; TELEFAX: (617) 542-8906

; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 411

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-286-748B-18

Query Match 100.0%; Score 44; DB 1; Length 411;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8

Db 136 KPSSPPE 143

## RESULT 15

US-08-153-799-18

; Sequence 18, Application US/08153799

; Patent No. 576683

; GENERAL INFORMATION:

; APPLICANT: Ballance, David J

; APPLICANT: Goodey, Andrew R

; TITLE OF INVENTION: Polypeptides

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: R Hain Swope, BOC Health Care Inc

; STREET: 100 Mountain Avenue

; CITY: Murray Hill

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07974

; COMPUTER READABLE FORM:

; MEDIUM TYPE: floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/153,799

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/847975

; FILING DATE: 06-MAR-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 8909916.2

; FILING DATE: 29-APR-1989

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/GB90/00650

; FILING DATE: 26-APR-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/775952

; FILING DATE: 29-OCT-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Swope, R Hain

; REGISTRATION NUMBER: 24864

; REFERENCE/DOCKET NUMBER: 92H832

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (908) 665 2400

; TELEFAX: (908) 771 6159

; TELEX: 219484

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 411 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-153-799-18

Query Match 100.0%; Score 44; DB 1; Length 411;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8

Db 136 KPSSPPE 143

## RESULT 16

US-08-560-098A-48

; Sequence 48, Application US/08560098A

; Patent No. 5976841

; GENERAL INFORMATION:

; APPLICANT: WENDET, Stephan

; APPLICANT: HEINZEL-WIELAND, Regina

; APPLICANT: STEFFENS, Gerd Josef

; TITLE OF INVENTION: Proteins having Fibrinolytic and

; TITLE OF INVENTION: Coagulation-inhibiting Properties

; NUMBER OF SEQUENCES: 60

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Evenson, McKeown, Edwards &amp; Lenahan

; STREET: 1200 G Street, N.W., Suite 700

; CITY: Washington

; STATE: DC

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/560,098A

; FILING DATE: 17-NOV-1995

; PRIOR APPLICATION DATA:

APPLICATION NUMBER: P 44 40 892.7  
FILING DATE: 17-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: EVANS, Joseph D.  
REGISTRATION NUMBER: 26,269  
REFERENCE/DOCKET NUMBER: 148/42448  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-8800  
TELEFAX: (202) 628-8844  
INFORMATION FOR SEQ ID NO: 48:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 411 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-560-098A-48

Query Match 100.0%; Score 44; DB 2; Length 411;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
Db 136 KPSSPPE 143

RESULT 17  
US-09-181-816-1  
Sequence 1, Application US/09181816  
Patent No. 6277818  
GENERAL INFORMATION:  
APPLICANT: MAZAR, Andrew P.  
APPLICANT: JONES, Terence R.  
TITLE OF INVENTION: CYCLIC PEPTIDE LIGANDS THAT TARGET UROKINASE  
FILE REFERENCE: 32904200300 SIDN 1-7  
CURRENT APPLICATION NUMBER: US/09/181,816  
CURRENT FILING DATE: 1998-10-29  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 1  
LENGTH: 411  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-181-816-1

Query Match 100.0%; Score 44; DB 3; Length 411;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
Db 136 KPSSPPE 143

RESULT 18  
US-09-403-736-2  
Sequence 2, Application US/09403736  
Patent No. 6638502  
GENERAL INFORMATION:  
APPLICANT: Aventis S.A.  
APPLICANT: Li, Hong  
APPLICANT: GRISCELLI, Frank  
APPLICANT: OPOLOU, Paule  
APPLICANT: SORIA, Claudine  
APPLICANT: RAGOT, Thierry  
APPLICANT: LEGRAND, Yves  
APPLICANT: SORIA, Jeanette  
APPLICANT: MABILAT, Christelle  
APPLICANT: PERRICAUDET, Michel  
APPLICANT: YEH, Patrice

TITLE OF INVENTION: Adenovirus-Mediated Intratumoral Delivery Of An Angiogenesis Antag  
TITLE OF INVENTION: For The Treatment Of Tumors  
FILE REFERENCE: A2778A-US  
CURRENT APPLICATION NUMBER: US/09/403,736  
CURRENT FILING DATE: 1999-10-26  
PRIOR APPLICATION NUMBER: PCT/EP98/02491  
PRIOR FILING DATE: 1998-11-05  
PRIOR APPLICATION NUMBER: 60/044,980  
PRIOR FILING DATE: 1997-04-28  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 2  
LENGTH: 411  
TYPE: PRT  
ORGANISM: humanurokinase  
US-09-403-736-2

Query Match 100.0%; Score 44; DB 4; Length 411;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
Db 136 KPSSPPE 143

RESULT 19  
US-07-942-157A-3  
Sequence 3, Application US/07942157A  
Patent No. 5648253  
GENERAL INFORMATION:  
APPLICANT: Wei, Cha-Mer  
TITLE OF INVENTION: Inhibitor-Resistant Urokinase  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kilpatrick & Cody  
STREET: 1100 Peachtree Street Suite 2800  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30309-4530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/942,157A  
FILING DATE: 19920908  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/631673  
FILING DATE: 20-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: TS1108Cont.  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404)815-6508  
TELEFAX: (404)815-6555  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 430 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..19  
OTHER INFORMATION: /label= peptide  
OTHER INFORMATION: /note= "MAP signal"  
FEATURE:

NAME/KEY: Modified-site  
LOCATION: 198..203  
OTHER INFORMATION: /label= modified  
OTHER INFORMATION: /note= "six amino acids deleted in mutant"  
US-07-942-157A-3

Query Match 100.0%; Score 44; DB 1; Length 430;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8  
Db 155 KPSSPPEE 162

RESULT 20  
5219569-2  
; Patent No. 5219569  
; APPLICANT: BLABER, MICHAEL; HEYNEKER, HERBERT L.; YEHAR,  
; GORDON A.  
; TITLE OF INVENTION: PROTEASE RESISTANT UROKINASE  
; NUMBER OF SEQUENCES: 6  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/766,858  
; FILING DATE: 16-AUG-1985  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 725,468  
; FILING DATE: 22-APR-1985  
; SEQ ID NO: 2  
; LENGTH: 430  
5219569-2

Query Match 100.0%; Score 44; DB 6; Length 430;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8  
Db 155 KPSSPPEE 162

RESULT 21  
US-09-101-272G-1  
; Sequence 1, Application US/09101272G  
; Patent No. 6509445  
; GENERAL INFORMATION:  
; APPLICANT: Nissin Food Products Co., Ltd.  
; TITLE OF INVENTION: CANCEROUS METASTASIS INHIBITOR  
; FILE REFERENCE: 050979  
; CURRENT APPLICATION NUMBER: US/09/101,272G  
; CURRENT FILING DATE: 1998-07-08  
; PRIOR APPLICATION NUMBER: JP 1059/1996  
; PRIOR FILING DATE: 1996-01-08  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: Patent version 3.1  
; SEQ ID NO: 1  
; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: mat peptide  
; LOCATION: (21)..()  
; OTHER INFORMATION:  
; NAME/KEY: misc\_feature  
; LOCATION: (20)..()  
; OTHER INFORMATION: Urokinase-type plasminogen activator (uPA)  
US-09-101-272G-1

Query Match 100.0%; Score 44; DB 4; Length 431;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8

Db 156 KPSSPPEE 163

RESULT 22  
5188829-1  
; Patent No. 5188829  
; APPLICANT: KOBAYASHI, YO-ICHI; OMORI, MUNEKI; YAMADA, CHIKAKO  
; TITLE OF INVENTION: RAPIDLY ACTING PROUROKINASE  
; NUMBER OF SEQUENCES: 23  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/340,007  
; FILING DATE: 18-AUG-1988  
; SEQ ID NO: 1  
; LENGTH: 431  
5188829-1

Query Match 100.0%; Score 44; DB 6; Length 431;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8  
Db 156 KPSSPPEE 163

RESULT 23  
US-08-560-098A-47  
; Sequence 47, Application US/08560098A  
; Patent No. 5976841  
; GENERAL INFORMATION:  
; APPLICANT: MNENDT, Stephan  
; APPLICANT: HEINZEL-WIELAND, Regina  
; APPLICANT: STEFFENS, Gerd Josef  
; TITLE OF INVENTION: Proteins having Fibrinolytic and  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Evenson, McKeown, Edwards & Lenahan  
; STREET: 1200 G Street, N.W., Suite 700  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/560,098A  
; FILING DATE: 17-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: P 44 40 892.7  
; FILING DATE: 17-NOV-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: EVANS, Joseph D.  
; REGISTRATION NUMBER: 26,269  
; REFERENCE/DOCKET NUMBER: 148/42448  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 628-8800  
; TELEFAX: (202) 628-8844  
; INFORMATION FOR SEQ ID NO: 47:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 432 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-560-098A-47

Query Match 100.0%; Score 44; DB 2; Length 432;  
Best Local Similarity 100.0%; Pred. No. 24;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KPSSPPEE 8  
|:|||||  
Db 157 KPSSPPEE 164

RESULT 24  
US-09-122-079-1  
; Sequence 1, Application US/09122079  
; Patent No. 6288024  
; GENERAL INFORMATION:  
; APPLICANT: Bouck, No. 62880241 P.  
; APPLICANT: Dawson, David W.  
; APPLICANT: Gillis, Paul R.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING ANGIOGENESIS  
; FILE REFERENCE: 87292  
; CURRENT APPLICATION NUMBER: US/09/122,079  
; PRIOR FILING DATE: 1998-07-23  
; EARLIER APPLICATION NUMBER: US 08/899304  
; PRIOR FILING DATE: 1997-07-23  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 362  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-122-079-1

Query Match 81.8%; Score 36; DB 3; Length 362;  
Best Local Similarity 85.7%; Pred. No. 3e+02; Indels 0; Gaps 0;  
Matches 6; Conservative 1; Mismatches 0;

QY 2 PSSPPEE 8  
|:|||||  
Db 22 PASPPEE 28

RESULT 25  
US-09-875-223-1  
; Sequence 1, Application US/09875223  
; Patent No. 6391850  
; GENERAL INFORMATION:  
; APPLICANT: No. 6391850thwestern University  
; APPLICANT: No. 63918501 Bouck  
; APPLICANT: David Dawson  
; APPLICANT: Paul Gillis  
; TITLE OF INVENTION: Methods and Compositions for Inhibiting Angiogenesis  
; FILE REFERENCE: 0290-2303  
; CURRENT APPLICATION NUMBER: US/09/875,223  
; PRIOR FILING DATE: 2001-06-06  
; PRIOR APPLICATION NUMBER: US 09/122,079  
; PRIOR FILING DATE: 1998-07-23  
; PRIOR APPLICATION NUMBER: PCT/US98/15228  
; PRIOR FILING DATE: 1998-07-23  
; PRIOR APPLICATION NUMBER: US 08/899,304  
; PRIOR FILING DATE: 1997-07-23  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 362  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-875-223-1

Query Match 81.8%; Score 36; DB 4; Length 362;  
Best Local Similarity 85.7%; Pred. No. 3e+02; Indels 0; Gaps 0;  
Matches 6; Conservative 1; Mismatches 0;

QY 2 PSSPPEE 8  
|:|||||  
Db 22 PASPPEE 28

RESULT 26  
US-09-875-114-1  
; Sequence 1, Application US/09875114  
; Patent No. 6670333  
; GENERAL INFORMATION:  
; APPLICANT: No. 6670333thwestern University  
; APPLICANT: No. 66703331 Bouck  
; APPLICANT: David Dawson  
; APPLICANT: Paul Gillis  
; TITLE OF INVENTION: Methods and Compositions for Inhibiting Angiogenesis  
; FILE REFERENCE: 0290-2302  
; CURRENT APPLICATION NUMBER: US/09/875,114  
; PRIOR FILING DATE: 2001-06-06  
; PRIOR APPLICATION NUMBER: US 09/122,079  
; PRIOR FILING DATE: 1998-07-23  
; PRIOR APPLICATION NUMBER: PCT/US98/15228  
; PRIOR FILING DATE: 1998-07-23  
; PRIOR APPLICATION NUMBER: US 08/899,304  
; PRIOR FILING DATE: 1997-07-23  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 362  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-875-114-1

Query Match 81.8%; Score 36; DB 4; Length 362;  
Best Local Similarity 85.7%; Pred. No. 3e+02; Indels 0; Gaps 0;  
Matches 6; Conservative 1; Mismatches 0;

QY 2 PSSPPEE 8  
|:|||||  
Db 22 PASPPEE 28

RESULT 27  
US-08-257-963B-2  
; Sequence 2, Application US/08257963B  
; Patent No. 5840686  
; GENERAL INFORMATION:  
; APPLICANT: Chader, Gerald J.; Becerra, S.  
; APPLICANT: Patricia; Schwartz, Joan P.;  
; APPLICANT: Taniwaki, Takayuki  
; TITLE OF INVENTION: PIGMENT EPITHELIUM  
; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION OF ITS NOVEL  
; TITLE OF INVENTION: BIOLOGICAL ACTIVITY AND SEQUENCES ENCODING  
; TITLE OF INVENTION: AND EXPRESSING THE PROTEIN  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Morgan & Finnegan  
; STREET: 345 Park Avenue  
; City: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC Compatible  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/257,963B  
; FILING DATE:  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/952,796  
; FILING DATE: 24-SEPT-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: DOROTHY R. AUTH  
; REGISTRATION NUMBER: 36434  
; REFERENCE/DOCKET NUMBER: 20264126US1



```
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 418 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1373
OTHER INFORMATION: /note="product =
OTHER INFORMATION: "pigment epithelial-derived factor"
OTHER INFORMATION: gene = "PEDF" codon_start = 1"
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 117..170
OTHER INFORMATION: PEDF amino acid
OTHER INFORMATION: sequence
US-08-257-963B-2

Query Match
Best Local Similarity 81.8%; Score 36; DB 2; Length 418;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PSSPPE 8
Db 22 PASPPE 28

RESULT 28
US-08-367-841A-2
Sequence 2, Application US/08367841A
Patent No. 631967
GENERAL INFORMATION:
APPLICANT: Chader, Gerald J.; Rodriguez,
APPLICANT: Ignacio R.; Mazuruk, Krzysztof;
APPLICANT: Tombran-Tink, Joyce
TITLE OF INVENTION: PIGMENT EPITHELIUM
TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSER: Morgan & Finnegan
STREET: 345 Park Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/367,841A
FILING DATE: 30-DEC-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/257,963
FILING DATE: 07-JUN-1994
PRIOR APPLICATION DATA: 07/952,796
FILING DATE: 24-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: DOROTHY R. AUTH
REGISTRATION NUMBER: 36434
REFERENCE/DOCKET NUMBER: 20264126US2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 751-6849
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 2:
```

```
SEQUENCE CHARACTERISTICS:
LENGTH: 418 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1373
OTHER INFORMATION: /note="product =
OTHER INFORMATION: "pigment epithelial-derived factor"
OTHER INFORMATION: gene = "PEDF" codon_start = 1"
FEATURE:
NAME/KEY:
LOCATION:
OTHER INFORMATION: PEDF amino acid
OTHER INFORMATION: sequence
US-08-367-841A-2

Query Match
Best Local Similarity 81.8%; Score 36; DB 4; Length 418;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PSSPPE 8
Db 22 PASPPE 28

RESULT 29
US-08-520-373D-2
Sequence 2, Application US/08520373D
Patent No. 6451763
GENERAL INFORMATION:
APPLICANT: Tombran-Tink, Joyce
APPLICANT: Steele, Finlan R
APPLICANT: Chader, Gerald J
APPLICANT: Becerra, Sofia P
APPLICANT: Johnson, Lincoln V
APPLICANT: Rodriguez, Ignacio R
TITLE OF INVENTION: RETINAL PIGMENTED EPITHELIUM DERIVED NEUROTROPIC FACTOR
FILE REFERENCE: 2026-4203US1
CURRENT APPLICATION NUMBER: US/08/520,373D
CURRENT FILING DATE: 1995-08-29
PRIOR APPLICATION NUMBER: 08/377,710
PRIOR FILING DATE: 1995-01-25
PRIOR APPLICATION NUMBER: 08/279,979
PRIOR FILING DATE: 1994-07-25
PRIOR APPLICATION NUMBER: 07/894,215
PRIOR FILING DATE: 1992-06-04
PRIOR APPLICATION NUMBER: 07/952,796
PRIOR FILING DATE: 1992-09-24
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 418
TYPE: PRT
ORGANISM: HUMAN
US-08-520-373D-2

Query Match
Best Local Similarity 81.8%; Score 36; DB 4; Length 418;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PSSPPE 8
Db 22 PASPPE 28

RESULT 30
PCT-US95-07201-2
Sequence 2, Application PC/TUS9507201
GENERAL INFORMATION:
APPLICANT: Chader, Gerald J.; Becerra, Sofia
APPLICANT: Patricia, Schwartz, Joan P.;
```

APPLICANT: Taniwaki, Takayuki  
TITLE OF INVENTION: PIGMENT EPITHELIAL  
TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC  
TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morgan & Finnegan, L.L.P.  
STREET: 345 Park Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/07201  
FILING DATE: 06-JUN-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/367,841  
FILING DATE: 30-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/257,963  
FILING DATE: 07-JUN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/952,796  
FILING DATE: 24-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: DOROTHY R. AUTH  
REGISTRATION NUMBER: 36434  
REFERENCE/DOCKET NUMBER: 20264126PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 418 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 117..1373  
OTHER INFORMATION: /note= "product =  
OTHER INFORMATION: "pigment epithelial-derived factor"  
OTHER INFORMATION: gene = "PEDF" codon\_start = 1"  
FEATURE:  
NAME/KEY:  
LOCATION:  
OTHER INFORMATION: PEDF amino acid  
OTHER INFORMATION: sequence  
PCT-US95-07201-2

Query Match 81.8%; Score 36; DB 5; Length 418;  
Best Local Similarity 85.7%; Pred. No. 3.5e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PSSPPEE 8  
DB 22 PPSPPPEE 28

RESULT 31  
US-08-291-349A-1  
Sequence 1, Application US/08291349A  
Patent No. 5545620  
GENERAL INFORMATION:  
APPLICANT: Wahl, Sharon M.  
APPLICANT: McCarthy, James B.  
APPLICANT: Furcht, Leo T.

TITLE OF INVENTION: Synthetic Fibronectin Fragments As  
TITLE OF INVENTION: Inhibitors of Retroviral Infection  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Allegretti & Witcoff  
STREET: 10 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/291,349A  
FILING DATE: 16 AUG 1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/006,121  
FILING DATE: 19 JAN 1993  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: McDonnell, John J.  
REGISTRATION NUMBER: 26,949  
REFERENCE/DOCKET NUMBER: 92,673-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1000  
TELEFAX: 312-715-1234  
TELEX: 910-221-5317  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..19  
OTHER INFORMATION: /note= "FNI (1906-1924)  
OTHER INFORMATION: fibronectin fragment : FN-C/H-I"  
US-08-291-349A-1

Query Match 79.5%; Score 35; DB 1; Length 19;  
Best Local Similarity 75.0%; Pred. No. 24;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8  
DB 3 KPSSPPEE 10

RESULT 32  
US-07-990-296-1  
Sequence 1, Application US/07990296  
Patent No. 5591719  
GENERAL INFORMATION:  
APPLICANT: Furcht, Leo T.  
APPLICANT: Allen, Janice B.  
APPLICANT: Wahl, Sharon M.  
APPLICANT: McCarthy, James B.  
TITLE OF INVENTION: Method for Treating Acute and  
TITLE OF INVENTION: Chronic Inflammatory Disorders using Polypeptides  
TITLE OF INVENTION: with Fibronectin Activity  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merchant & Gould  
STREET: 3100 No. 5591719west Center  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA

ZIP: 55402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Wordperfect 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/990,296  
FILING DATE: 19921210  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Kowalchuk, Alan W.  
REGISTRATION NUMBER: 31,535  
REFERENCE/DOCKET NUMBER: 600,252-US-01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-332-5300  
TELEFAX: 612-332-9081  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19 amino acid residues  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: Internal Fragment  
ORIGINAL SOURCE: Synthetically Derived  
FEATURE:  
NAME/KEY: Fragment of the 33 kD carboxy  
terminal heparin-binding fragment of the A  
NAME/KEY: Chain of fibronectin  
LOCATION: Represents isolated fibronectin  
LOCATION: residues 1906-1924 from all plasma isoforms  
LOCATION: of fibronectin  
US-07-990-296-1

Query Match 79.5%; Score 35; DB 1; Length 19;  
Best Local Similarity 75.0%; Pred. No. 24;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KPSSPPE 8  
Db 3 KPSSPPE 10

RESULT 33  
US-08-480-133A-1  
Sequence 1, Application US/08460133A  
Patent No. 5840691  
GENERAL INFORMATION:  
APPLICANT: Furcht, Leo T.  
APPLICANT: McCarthy, James B.  
APPLICANT: Wahl, Sharon M.  
APPLICANT: Allen, Janice B.  
APPLICANT: Billups, Kevin L.  
APPLICANT: Everett, Jeffrey E.  
TITLE OF INVENTION: Method for Treating Inflammatory  
Diseases Using Polypeptides with Fibronectin Activity  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merchant & Gould  
STREET: 90 South 7th Street, 3100 No. 5840691west Center  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,133A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/139,903  
FILING DATE: 21-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/990,296  
FILING DATE: 10-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Carter, Charles G.  
REGISTRATION NUMBER: 35,093  
REFERENCE/DOCKET NUMBER: 600-308US01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-332-5300  
TELEFAX: 612-332-9081  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-480-133A-1

Query Match 79.5%; Score 35; DB 2; Length 19;  
Best Local Similarity 75.0%; Pred. No. 24;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KPSSPPE 8  
Db 3 KPSSPPE 10

RESULT 34  
US-08-394-748A-9  
Sequence 9, Application US/08394748A  
Patent No. 6013628  
GENERAL INFORMATION:  
APPLICANT: Skubitz, Amy P.N.  
APPLICANT: Furcht, Leo T.  
APPLICANT: Bales, Mark  
APPLICANT: Gregerson, Dale S.  
APPLICANT: Agarwal, Anita  
APPLICANT: Wright, Martha M.  
APPLICANT: Murali, Shobana  
TITLE OF INVENTION: Method for Treating Conditions of the Eye  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merchant & Gould  
STREET: 3100 No. 6013628west Center  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/394,748A  
FILING DATE: 27-FEB-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/203,458  
FILING DATE: 28-FEB-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Carter, Charles G.  
REGISTRATION NUMBER: 35,093  
REFERENCE/DOCKET NUMBER: 600.307US01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-332-5300  
TELEFAX: 612-332-9081  
INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:  
LENGTH: 19 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE: FN-C/H-1  
STRAIN: FN-C/H-1  
US-08-394-748A-9

Query Match 79.5%; Score 35; DB 3; Length 19;  
Best Local Similarity 75.0%; Pred. No. 24;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
Db 3 KPSSPPE 10

RESULT 35  
US-09-675-922-13  
Sequence 13, Application US/09675922  
Patent No. 6468731  
GENERAL INFORMATION:  
APPLICANT: Hubbell A., Jeffrey  
APPLICANT: Schense C., Jason  
APPLICANT: Sakiyama E., Shelley  
TITLE OF INVENTION: Enzyme-Mediated Modification of Fibrin for Tissue  
TITLE OF INVENTION: Engineering: Incorporation of Proteins  
FILE REFERENCE: ETH 107 DIV  
CURRENT APPLICATION NUMBER: US/09/675,922  
CURRENT FILING DATE: 2000-09-29  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 13  
LENGTH: 19  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:  
OTHER INFORMATION: heparin-binding sequence  
US-09-675-922-13

Query Match 79.5%; Score 35; DB 4; Length 19;  
Best Local Similarity 75.0%; Pred. No. 24;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
Db 3 KPSSPPE 10

RESULT 36  
PCT-US93-11781-1  
Sequence 1, Application PC/TUS9311781  
GENERAL INFORMATION:  
APPLICANT: Regents of the University of Minnesota  
APPLICANT: Morrill Hall  
APPLICANT: 100 Church Street, S.E.  
APPLICANT: Minneapolis, Minnesota 55455  
APPLICANT: U.S.A.  
APPLICANT: Represented By The Secretary of Health  
APPLICANT: and Human Services  
APPLICANT: 200 Independence Avenue S.W.  
APPLICANT: Washington, D.C. 20201  
APPLICANT: United States of America  
TITLE OF INVENTION: Polypeptides Useful for Treating  
TITLE OF INVENTION: Inflammatory Disorders  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merchant & Gould  
STREET: 3100 Norwest Center  
CITY: Minneapolis  
STATE: MN

COUNTRY: USA  
ZIP: 55402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/11781  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/139,903  
FILING DATE: 21-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/990,296  
FILING DATE: 10-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Kowalczyk, Alan W.  
REGISTRATION NUMBER: 31,535  
REFERENCE/DOCKET NUMBER: 600.283-US-01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-332-5300  
TELEFAX: 612-332-9081  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US93-11781-1

Query Match 79.5%; Score 35; DB 5; Length 19;  
Best Local Similarity 75.0%; Pred. No. 24;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
Db 3 KPSSPPE 10

RESULT 37  
PCT-US95-02478-9  
Sequence 9, Application PC/TUS9502478  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: Method for Treating Conditions  
TITLE OF INVENTION: Of the Eye Using Polypeptides  
NUMBER OF SEQUENCES: 16  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/02478  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/203,458  
FILING DATE: 28-FEB-1994  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
STRAIN: FN-C/H-1  
PCT-US95-02478-9

Query Match 79.5%; Score 35; DB 5; Length 19;  
Best Local Similarity 75.0%; Pred. No. 24;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KPSPPE 8  
|||  
Db 3 KPSPPE 10

## RESULT 38

US-09-252-991A-29265

Sequence 29265, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfeld et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 29265

LENGTH: 150

TYPE: PR

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-29265

Query Match 79.5%; Score 35; DB 4; Length 150;  
Best Local Similarity 75.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KPSPPE 8  
|||  
Db 104 KPSPPE 111

## RESULT 39

US-08-536-891A-1

Sequence 1, Application US/08536891A

Patent No. 6033907

GENERAL INFORMATION:

APPLICANT: David A. Williams

TITLE OF INVENTION: Enhanced Virus-Mediated DNA Transfer

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS: Thomas Q. Henry

STREET: Bank One Tower, Suite 3700, 111 Monument Circle

CITY: Indianapolis

STATE: Indiana

COUNTRY: USA

ZIP: 46204

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage

COMPUTER: COMPAQ

OPERATING SYSTEM: MSDOS

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/536,891A

FILING DATE: September 29, 1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/03817

FILING DATE: March 27, 1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/218,355

FILING DATE: March 25, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Thomas Q. Henry

REGISTRATION NUMBER: 28,309

REFERENCE/DOCKET NUMBER: 7037-52/1U-33-CIP-2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (317) 634-3456  
TELEFAX: (317) 637-7561  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 271 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein fragment

## US-08-536-891A-1

Query Match

Best Local Similarity

Matches 6; Conservative

QY 1 KPSPPE 8  
|||  
Db 219 KPSPPE 226

## RESULT 40

US-08-933-100B-14

Sequence 14, Application US/08933100B

Patent No. 6274704

GENERAL INFORMATION:

APPLICANT: FUKAI, FUMIO

TITLE OF INVENTION: BIOLOGICALLY ACTIVE PEPTIDE AND CANCER

TITLE OF INVENTION: METASTASIS INHIBITOR

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSER: PILLSBURY, MADISON &amp; SUTRO

STREET: 1100 NEW YORK AVENUE, N.W.

CITY: WASHINGTON

STATE: D.C.

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/933,100B

FILING DATE: 18-SEP-1997

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: PERRY, GLENN

REGISTRATION NUMBER: 28458

REFERENCE/DOCKET NUMBER: 7898/242094

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-861-3000

TELEFAX: 202-822-0944

TELEX: 6714627 CUSH

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 271

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

OTHER INFORMATION: sequence of a part (1600-1870) of Heparin binding site

US-08-933-100B-14

Query Match

Best Local Similarity

Matches 6; Conservative

QY 1 KPSPPE 8  
|||  
Db 219 KPSPPE 226

RESULT 41  
US-09-366-009-1  
; Sequence 1, Application US/09366009  
; Patent No. 6426042  
; GENERAL INFORMATION:  
; APPLICANT: Asada, Kiyozo  
; Uemori, Takashi  
; Koyama, No. 6426042uto  
; Hashino, Kimikazu  
; Kato, Ikunoshin  
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET  
; CELLS WITH RETROVIRUS  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: WEISER & ASSOCIATES  
; STREET: 230 South Fifteenth Street, Suite 500  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19102  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/366,009  
; FILING DATE: 02-Aug-1999  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/809,156  
; FILING DATE: <Unknown>  
; APPLICATION NUMBER: JP 294382/1995  
; FILING DATE: 13-NOV-1995  
; APPLICATION NUMBER: JP 051847/1996  
; FILING DATE: 08-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weiser, Gerard J.  
; REGISTRATION NUMBER: 19,763  
; REFERENCE/DOCKET NUMBER: 977.6507P  
; TELEPHONE: 215-875-8383  
; TELEFAX: 215-875-8394  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 271 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: <Unknown>  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-366-009-1

Query Match 79.5%; Score 35; DB 4; Length 271;  
Best Local Similarity 75.0%; Pred. No. 3.2e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
Db 219 KPSSPPE 226

RESULT 42  
US-08-809-156B-1  
; Sequence 1, Application US/08809156B  
; Patent No. 6472204  
; GENERAL INFORMATION:  
; APPLICANT: Asada, Kiyozo  
; Uemori, Takashi  
; Koyama, No. 6472204uto  
; Hashino, Kimikazu

APPLICANT: Kato, Ikunoshin  
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET  
; CELLS WITH RETROVIRUS  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: WEISER & ASSOCIATES  
; STREET: 230 South Fifteenth Street, Suite 500  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19102  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/809,156B  
; FILING DATE: 07-MAR-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP96/03254  
; FILING DATE: 07-NOV-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 294382/1995  
; FILING DATE: 13-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 051847/1996  
; FILING DATE: 08-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weiser, Gerard J.  
; REGISTRATION NUMBER: 19,763  
; REFERENCE/DOCKET NUMBER: 977.6507P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-875-8383  
; TELEFAX: 215-875-8394  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 271 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-809-156B-1

Query Match 79.5%; Score 35; DB 4; Length 271;  
Best Local Similarity 75.0%; Pred. No. 3.2e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
Db 219 KPSSPPE 226

RESULT 43  
US-09-043-981-1  
; Sequence 1, Application US/09043981  
; Patent No. 6670177  
; GENERAL INFORMATION:  
; APPLICANT: Williams, David A.  
; TITLE OF INVENTION: METHODS FOR ENHANCED VIRUS MEDIATED DNA TRANSFER USING  
; FILE REFERENCE: 7037-297  
; CURRENT APPLICATION NUMBER: US/09/043,981  
; FILING DATE: 1998-06-19  
; EARLIER APPLICATION NUMBER: US96/15712  
; EARLIER FILING DATE: 1996-09-30  
; EARLIER APPLICATION NUMBER: 08/536,891  
; EARLIER FILING DATE: 1995-09-29  
; EARLIER APPLICATION NUMBER: 60/024,169  
; EARLIER FILING DATE: 1996-08-19  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 1  
LENGTH: 271  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-043-981-1

Query Match 79.5%; Score 35; DB 4; Length 271;  
Best Local Similarity 75.0%; Pred. No. 3.2e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KPSPPEE 8  
DB 219 KPSPPEE 226

RESULT 44  
US-08-836-854-4  
Sequence 4, Application US/08836854  
Patent No. 5824547  
GENERAL INFORMATION:  
APPLICANT: HASHINO, Kimikazu  
APPLICANT: MATSUSHITA, Hideyuki  
APPLICANT: KATO, Ikunoshin  
TITLE OF INVENTION: METHOD OF PRODUCTION OF TRANSFECTED CELLS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street N.W. Ste. 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/836,854  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP95/02425  
FILING DATE: 29-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 317721/1994  
FILING DATE: 29-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: HASHINO=1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 296 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-836-854-4

Query Match 79.5%; Score 35; DB 2; Length 296;  
Best Local Similarity 75.0%; Pred. No. 3.5e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KPSPPEE 8  
DB 219 KPSPPEE 226

RESULT 45

US-08-836-854-17  
Sequence 17, Application US/08836854  
Patent No. 5824547  
GENERAL INFORMATION:  
APPLICANT: HASHINO, Kimikazu  
APPLICANT: MATSUSHITA, Hideyuki  
APPLICANT: KATO, Ikunoshin  
TITLE OF INVENTION: METHOD OF PRODUCTION OF TRANSFECTED CELLS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street N.W. Ste. 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/836,854  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP95/02425  
FILING DATE: 29-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 317721/1994  
FILING DATE: 29-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: HASHINO=1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 368 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-836-854-17

Query Match 79.5%; Score 35; DB 2; Length 368;  
Best Local Similarity 75.0%; Pred. No. 4.3e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KPSPPEE 8  
DB 316 KPSPPEE 323

RESULT 46  
US-08-836-854-16  
Sequence 16, Application US/08836854  
Patent No. 5824547  
GENERAL INFORMATION:  
APPLICANT: HASHINO, Kimikazu  
APPLICANT: MATSUSHITA, Hideyuki  
APPLICANT: KATO, Ikunoshin  
TITLE OF INVENTION: METHOD OF PRODUCTION OF TRANSFECTED CELLS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street N.W. Ste. 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/836,854  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP95/02425  
FILING DATE: 29-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 317721/1994  
FILING DATE: 29-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: HASHINO-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 457 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-836-854-16

Query Match 79.5%; Score 35; DB 2; Length 457;  
Best Local Similarity 75.0%; Pred. No. 5.3e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 KPSSPPE 8  
DB 405 KPSSPPE 412

RESULT 47  
US-09-366-009-22  
Sequence 22, Application US/09366009  
Patent No. 6426042  
GENERAL INFORMATION:  
APPLICANT: Asada, Kiyozo  
Uemori, Takashi  
Ueno, Takashi  
Koyama, No. 6426042uto  
Hashino, Kimikazu  
Kato, Ikunoshin  
TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET  
CELLS WITH RETROVIRUS  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: WEISER & ASSOCIATES  
STREET: 230 South Fifteenth Street, Suite 500  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/366,009  
FILING DATE: 02-AUG-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/809,156  
FILING DATE: <Unknown>

APPLICATION NUMBER: JP 294382/1995  
FILING DATE: 13-NOV-1995  
APPLICATION NUMBER: JP 051847/1996  
FILING DATE: 08-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Weiser, Gerard J.  
REGISTRATION NUMBER: 19,763  
REFERENCE/DOCKET NUMBER: 977.6507P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-875-8383  
TELEFAX: 215-875-8394  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 457 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 22:  
US-09-366-009-22

Query Match 79.5%; Score 35; DB 4; Length 457;  
Best Local Similarity 75.0%; Pred. No. 5.3e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 KPSSPPE 8  
DB 405 KPSSPPE 412

RESULT 48  
US-08-809-156B-22  
Sequence 22, Application US/08809156B  
Patent No. 6472204  
GENERAL INFORMATION:  
APPLICANT: Asada, Kiyozo  
Uemori, Takashi  
Ueno, Takashi  
Koyama, No. 6472204uto  
Hashino, Kimikazu  
Kato, Ikunoshin  
TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET  
CELLS WITH RETROVIRUS  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: WEISER & ASSOCIATES  
STREET: 230 South Fifteenth Street, Suite 500  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/809,156B  
FILING DATE: 07-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP96/03254  
FILING DATE: 07-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 294382/1995  
FILING DATE: 13-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 051847/1996  
FILING DATE: 08-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Weiser, Gerard J.  
REGISTRATION NUMBER: 19,763  
REFERENCE/DOCKET NUMBER: 977.6507P



TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-875-8383  
TELEFAX: 215-875-8394  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 457 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-809-156B-22

Query Match 79.5%; Score 35; DB 4; Length 457;  
Best Local Similarity 75.0%; Pred. No. 5.3e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KPSPPEE 8  
DB 405 KPSPPRE 412

RESULT 49  
US-09-366-009-13  
Sequence 13, Application US/09366009  
Patent No. 6426042  
GENERAL INFORMATION:  
APPLICANT: Asada, Kiyozo  
Uemori, Takashi  
Koyama, No. 6426042uto  
Hashino, Kimikazu  
Kato, Ikunoshin  
TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET  
CELLS WITH RETROVIRUS  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: WEISER & ASSOCIATES  
STREET: 230 South Fifteenth Street, Suite 500  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/366,009  
FILING DATE: 02-Aug-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/809,156  
FILING DATE: <Unknown>  
APPLICATION NUMBER: JP 294382/1995  
FILING DATE: 13-NOV-1995  
APPLICATION NUMBER: JP 051847/1996  
FILING DATE: 08-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Weiser, Gerard J.  
REGISTRATION NUMBER: 19,763  
REFERENCE/DOCKET NUMBER: 977.6507P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-875-8383  
TELEFAX: 215-875-8394  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 547 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 13:

US-09-366-009-13

Query Match 79.5%; Score 35; DB 4; Length 547;  
Best Local Similarity 75.0%; Pred. No. 6.4e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KPSPPEE 8  
DB 222 KPSPPRE 229

RESULT 50  
US-08-809-156B-13  
Sequence 13, Application US/08809156B  
Patent No. 6472204  
GENERAL INFORMATION:  
APPLICANT: Asada, Kiyozo  
Uemori, Takashi  
Koyama, No. 6472204uto  
Hashino, Kimikazu  
Kato, Ikunoshin  
TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET  
CELLS WITH RETROVIRUS  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: WEISER & ASSOCIATES  
STREET: 230 South Fifteenth Street, Suite 500  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/809,156B  
FILING DATE: 07-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP96/03254  
FILING DATE: 07-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 294382/1995  
FILING DATE: 13-NOV-1995  
APPLICATION NUMBER: JP 051847/1996  
FILING DATE: 08-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Weiser, Gerard J.  
REGISTRATION NUMBER: 19,763  
REFERENCE/DOCKET NUMBER: 977.6507P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-875-8383  
TELEFAX: 215-875-8394  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 547 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-809-156B-13

Query Match 79.5%; Score 35; DB 4; Length 547;  
Best Local Similarity 75.0%; Pred. No. 6.4e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KPSPPEE 8  
DB 222 KPSPPRE 229

Search completed: June 18, 2004, 13:00:59  
Job time : 27 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2004, 12:58:49 ; Search time 43 Seconds  
(without alignments)  
52.523 Million cell updates/sec

Title: US-10-655-201-2

Perfect score: 44

Sequence: 1 KPSPPER 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 500 summaries

Database :

Published Applications AA.\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	100.0	8	14 US-10-235-552-2	Sequence 2, Appl1
2	44	100.0	10	14 US-10-235-552-1	Sequence 9, Appl1
3	44	100.0	96	9 US-09-880-503-9	Sequence 1, Appl1
4	44	100.0	143	9 US-09-880-503-8	Sequence 8, Appl1
5	44	100.0	276	9 US-09-880-503-5	Sequence 5, Appl1
6	44	100.0	323	9 US-09-880-503-7	Sequence 7, Appl1
7	44	100.0	337	15 US-10-106-698-6266	Sequence 6266, Ap
8	44	100.0	337	15 US-10-264-049-2927	Sequence 2927, Ap
9	44	100.0	411	9 US-09-880-503-3	Sequence 3, Appl1
10	44	100.0	411	15 US-10-407-821-2	Sequence 2, Appl1
11	44	100.0	411	16 US-10-372-966-6	Sequence 6, Appl1
12	44	100.0	430	16 US-10-372-966-4	Sequence 4, Appl1
13	44	100.0	431	9 US-09-264-4688-1	Sequence 1, Appl1
14	44	100.0	431	12 US-10-411-037-34	Sequence 34, Appl1
15	44	100.0	431	12 US-10-411-026-34	Sequence 34, Appl1

16	44	100.0	431	12 US-10-282-174-562	Sequence 562, App
17	44	100.0	431	13 US-10-076-421-2	Sequence 2, Appl1
18	44	100.0	431	14 US-10-171-311-184	Sequence 184, App
19	44	100.0	431	14 US-10-193-656-4	Sequence 4, Appl1
20	44	100.0	431	14 US-10-301-822-161	Sequence 161, App
21	44	100.0	431	14 US-10-247-671-149	Sequence 149, App
22	44	100.0	431	14 US-10-131-985-21	Sequence 21, Appl1
23	44	100.0	431	15 US-10-295-027-414	Sequence 414, App
24	44	100.0	431	15 US-10-295-027-1275	Sequence 1275, Ap
25	44	100.0	431	16 US-10-410-962-34	Sequence 34, Appl1
26	44	100.0	431	16 US-10-411-049-34	Sequence 34, Appl1
27	44	100.0	433	16 US-10-372-966-5	Sequence 5, Appl1
28	44	100.0	437	16 US-10-087-192-594	Sequence 594, App
29	44	100.0	445	15 US-10-360-101-266	Sequence 266, App
30	41	93.2	8	14 US-10-235-552-8	Sequence 8, Appl1
31	39	88.6	7	14 US-10-235-552-4	Sequence 4, Appl1
32	39	88.6	7	14 US-10-235-552-5	Sequence 5, Appl1
33	38	86.4	8	14 US-10-235-552-6	Sequence 6, Appl1
34	38	86.4	8	14 US-10-235-552-7	Sequence 7, Appl1
35	38	86.4	166	9 US-09-864-761-37883	Sequence 37883, A
36	38	86.4	166	9 US-10-108-260A-3871	Sequence 46504, A
37	38	86.4	264	15 US-10-176-306-23	Sequence 3871, Ap
38	38	86.4	461	14 US-10-094-749-2145	Sequence 23, Appl1
39	38	86.4	461	15 US-10-094-749-2145	Sequence 2145, Ap
40	38	86.4	461	15 US-10-295-027-205	Sequence 205, App
41	38	86.4	461	16 US-10-188-832-97	Sequence 97, Appl1
42	37	84.1	8	14 US-10-223-172A-20	Sequence 20, Appl1
43	37	84.1	8	14 US-10-224-999A-20	Sequence 20, Appl1
44	37	84.1	8	16 US-10-663-407-20	Sequence 20, Appl1
45	37	84.1	9	14 US-10-663-407-20	Sequence 24, Appl1
46	37	84.1	9	14 US-10-223-172A-24	Sequence 24, Appl1
47	37	84.1	9	14 US-10-224-999A-24	Sequence 24, Appl1
48	37	84.1	9	16 US-10-663-407-24	Sequence 24, Appl1
49	37	84.1	10	9 US-09-972-035A-26	Sequence 26, Appl1
50	37	84.1	10	9 US-10-663-407-26	Sequence 26, Appl1
51	37	84.1	10	14 US-10-223-172A-29	Sequence 29, Appl1
52	37	84.1	10	14 US-10-224-999A-29	Sequence 29, Appl1
53	37	84.1	10	16 US-10-663-407-26	Sequence 26, Appl1
54	37	84.1	13	9 US-09-972-035A-30	Sequence 30, Appl1
55	37	84.1	13	14 US-10-223-172A-37	Sequence 37, Appl1
56	37	84.1	13	14 US-10-224-999A-37	Sequence 37, Appl1
57	37	84.1	13	16 US-10-663-407-30	Sequence 30, Appl1
58	37	84.1	14	9 US-09-972-035A-32	Sequence 32, Appl1
59	37	84.1	14	16 US-10-663-407-32	Sequence 32, Appl1
60	37	84.1	1099	12 US-10-263-928-135	Sequence 135, App
61	36	81.8	9	14 US-10-223-172A-27	Sequence 27, Appl1
62	36	81.8	9	14 US-10-224-999A-27	Sequence 27, Appl1
63	36	81.8	10	14 US-10-223-172A-33	Sequence 33, Appl1
64	36	81.8	10	14 US-10-224-999A-33	Sequence 33, Appl1
65	36	81.8	81	12 US-10-424-599-26528	Sequence 26528, App
66	36	81.8	84	12 US-10-262-839-36	Sequence 36, Appl1
67	36	81.8	362	9 US-09-873-114-1	Sequence 1, Appl1
68	36	81.8	418	12 US-10-262-839-34	Sequence 34, Appl1
69	36	81.8	418	12 US-10-180-959-1	Sequence 1, Appl1
70	36	81.8	418	14 US-10-020-541-1	Sequence 2, Appl1
71	36	81.8	418	14 US-10-216-373-2	Sequence 1, Appl1
72	36	81.8	418	15 US-10-342-243-1	Sequence 1, Appl1
73	36	81.8	418	15 US-10-603-387-1	Sequence 1, Appl1
74	36	81.8	425	12 US-10-424-599-245111	Sequence 245111, Sequence 201859, A
75	36	81.8	449	12 US-10-424-599-201859	Sequence 58413, A
76	36	81.8	510	12 US-10-425-114-58413	Sequence 30, Appl1
77	36	81.8	671	15 US-10-321-802-30	Sequence 6, Appl1
78	36	81.8	747	9 US-09-874-069-6	Sequence 2175, Ap
79	36	81.8	851	16 US-10-408-765A-2175	Sequence 102, App
80	36	81.8	988	9 US-09-874-069-4	Sequence 266, App
81	36	81.8	1089	12 US-10-206-915-266	Sequence 266, App
82	36	81.8	1089	12 US-10-199-670-266	Sequence 266, App
83	36	81.8	1089	12 US-10-201-858-266	Sequence 266, App
84	36	81.8	1089	12 US-10-205-890-266	Sequence 266, App
85	36	81.8	1089	12 US-10-208-024-266	Sequence 266, App
86	36	81.8	1089	12 US-10-201-851-266	Sequence 266, App
87	36	81.8	1089	12 US-10-174-581-266	Sequence 266, App
88	36	81.8	1089	12 US-10-174-581-266	Sequence 266, App

89	36	81.8	1089	12	US-10-176-483-266	Sequence 266, App	162	36	81.8	1089	14	US-10-184-658-266	Sequence 266, App
90	36	81.8	1089	12	US-10-176-749-266	Sequence 266, App	163	36	81.8	1089	14	US-10-176-991-266	Sequence 266, App
91	36	81.8	1089	12	US-10-176-914-266	Sequence 266, App	164	36	81.8	1089	14	US-10-173-695-266	Sequence 266, App
92	36	81.8	1089	12	US-10-176-915-266	Sequence 266, App	165	36	81.8	1089	14	US-10-173-797-266	Sequence 266, App
93	36	81.8	1089	12	US-10-006-485A-102	Sequence 102, App	166	36	81.8	1089	14	US-10-173-705-266	Sequence 266, App
94	36	81.8	1089	12	US-10-013-907A-102	Sequence 102, App	167	36	81.8	1089	14	US-10-174-578-266	Sequence 266, App
95	36	81.8	1089	12	US-10-015-499A-102	Sequence 102, App	168	36	81.8	1089	14	US-10-174-588-266	Sequence 266, App
96	36	81.8	1089	12	US-10-176-484-266	Sequence 266, App	169	36	81.8	1089	14	US-10-174-588-266	Sequence 266, App
97	36	81.8	1089	12	US-10-180-550-266	Sequence 266, App	170	36	81.8	1089	14	US-10-175-747-266	Sequence 266, App
98	36	81.8	1089	12	US-10-183-014-266	Sequence 266, App	171	36	81.8	1089	14	US-10-176-481-266	Sequence 266, App
99	36	81.8	1089	12	US-10-187-738-266	Sequence 266, App	172	36	81.8	1089	14	US-10-176-488-266	Sequence 266, App
100	36	81.8	1089	12	US-10-187-740-266	Sequence 266, App	173	36	81.8	1089	14	US-10-176-487-266	Sequence 266, App
101	36	81.8	1089	12	US-10-187-883-266	Sequence 266, App	174	36	81.8	1089	14	US-10-176-493-266	Sequence 266, App
102	36	81.8	1089	12	US-10-194-363-266	Sequence 266, App	175	36	81.8	1089	14	US-10-176-756-266	Sequence 266, App
103	36	81.8	1089	12	US-10-194-460-266	Sequence 266, App	176	36	81.8	1089	14	US-10-176-911-266	Sequence 266, App
104	36	81.8	1089	12	US-10-194-463-266	Sequence 266, App	177	36	81.8	1089	14	US-10-176-925-266	Sequence 266, App
105	36	81.8	1089	12	US-10-194-484-266	Sequence 266, App	178	36	81.8	1089	14	US-10-176-925-266	Sequence 266, App
106	36	81.8	1089	12	US-10-195-884-266	Sequence 266, App	179	36	81.8	1089	14	US-10-176-978-266	Sequence 266, App
107	36	81.8	1089	12	US-10-195-896-266	Sequence 266, App	180	36	81.8	1089	14	US-10-179-510-266	Sequence 266, App
108	36	81.8	1089	12	US-10-196-744-266	Sequence 266, App	181	36	81.8	1089	14	US-10-180-543-266	Sequence 266, App
109	36	81.8	1089	12	US-10-196-755-266	Sequence 266, App	182	36	81.8	1089	14	US-10-180-544-266	Sequence 266, App
110	36	81.8	1089	12	US-10-196-757-266	Sequence 266, App	183	36	81.8	1089	14	US-10-180-546-266	Sequence 266, App
111	36	81.8	1089	12	US-10-197-710-266	Sequence 266, App	184	36	81.8	1089	14	US-10-180-547-266	Sequence 266, App
112	36	81.8	1089	12	US-10-197-710-266	Sequence 266, App	185	36	81.8	1089	14	US-10-180-548-266	Sequence 266, App
113	36	81.8	1089	12	US-10-198-758-266	Sequence 266, App	186	36	81.8	1089	14	US-10-180-553-266	Sequence 266, App
114	36	81.8	1089	12	US-10-198-766-266	Sequence 266, App	187	36	81.8	1089	14	US-10-180-559-266	Sequence 266, App
115	36	81.8	1089	12	US-10-199-304-266	Sequence 266, App	188	36	81.8	1089	14	US-10-181-000-266	Sequence 266, App
116	36	81.8	1089	12	US-10-199-309-266	Sequence 266, App	189	36	81.8	1089	14	US-10-183-012-266	Sequence 266, App
117	36	81.8	1089	12	US-10-199-313-266	Sequence 266, App	190	36	81.8	1089	14	US-10-184-654-266	Sequence 266, App
118	36	81.8	1089	12	US-10-199-456-266	Sequence 266, App	191	36	81.8	1089	14	US-10-184-654-266	Sequence 266, App
119	36	81.8	1089	12	US-10-201-329-266	Sequence 266, App	192	36	81.8	1089	14	US-10-184-652-266	Sequence 266, App
120	36	81.8	1089	12	US-10-202-412-266	Sequence 266, App	193	36	81.8	1089	14	US-10-184-635-266	Sequence 266, App
121	36	81.8	1089	12	US-10-206-919-266	Sequence 266, App	194	36	81.8	1089	14	US-10-184-637-266	Sequence 266, App
122	36	81.8	1089	12	US-10-206-922-266	Sequence 266, App	195	36	81.8	1089	14	US-10-184-646-266	Sequence 266, App
123	36	81.8	1089	12	US-10-206-924-266	Sequence 266, App	196	36	81.8	1089	14	US-10-184-652-266	Sequence 266, App
124	36	81.8	1089	12	US-10-206-928-266	Sequence 266, App	197	36	81.8	1089	14	US-10-184-652-266	Sequence 266, App
125	36	81.8	1089	12	US-10-207-914-266	Sequence 266, App	198	36	81.8	1089	14	US-10-187-594-266	Sequence 266, App
126	36	81.8	1089	12	US-10-207-921-266	Sequence 266, App	199	36	81.8	1089	14	US-10-187-596-266	Sequence 266, App
127	36	81.8	1089	12	US-10-207-922-266	Sequence 266, App	200	36	81.8	1089	14	US-10-187-745-266	Sequence 266, App
128	36	81.8	1089	12	US-10-208-027-266	Sequence 266, App	201	36	81.8	1089	14	US-10-187-885-266	Sequence 266, App
129	36	81.8	1089	12	US-10-013-910A-102	Sequence 102, App	202	36	81.8	1089	14	US-10-187-886-266	Sequence 266, App
130	36	81.8	1089	12	US-10-174-570-266	Sequence 266, App	203	36	81.8	1089	14	US-10-199-484-266	Sequence 266, App
131	36	81.8	1089	12	US-10-183-005-266	Sequence 266, App	204	36	81.8	1089	14	US-10-196-756-266	Sequence 266, App
132	36	81.8	1089	12	US-10-226-254A-102	Sequence 102, App	205	36	81.8	1089	14	US-10-176-751-266	Sequence 266, App
133	36	81.8	1089	12	US-10-015-395A-102	Sequence 102, App	206	36	81.8	1089	14	US-10-176-760-266	Sequence 266, App
134	36	81.8	1089	13	US-10-052-586-266	Sequence 266, App	207	36	81.8	1089	14	US-10-176-990-266	Sequence 266, App
135	36	81.8	1089	14	US-10-174-590-266	Sequence 266, App	208	36	81.8	1089	14	US-10-180-541-266	Sequence 266, App
136	36	81.8	1089	14	US-10-176-758-266	Sequence 266, App	209	36	81.8	1089	14	US-10-180-542-266	Sequence 266, App
137	36	81.8	1089	14	US-10-175-737-266	Sequence 266, App	210	36	81.8	1089	14	US-10-180-546-266	Sequence 266, App
138	36	81.8	1089	14	US-10-173-706-266	Sequence 266, App	211	36	81.8	1089	14	US-10-180-551-266	Sequence 266, App
139	36	81.8	1089	14	US-10-175-738-266	Sequence 266, App	212	36	81.8	1089	14	US-10-180-552-266	Sequence 266, App
140	36	81.8	1089	14	US-10-175-752-266	Sequence 266, App	213	36	81.8	1089	14	US-10-180-999-266	Sequence 266, App
141	36	81.8	1089	14	US-10-176-482-266	Sequence 266, App	214	36	81.8	1089	14	US-10-183-013-266	Sequence 266, App
142	36	81.8	1089	14	US-10-176-757-266	Sequence 266, App	215	36	81.8	1089	14	US-10-184-612-266	Sequence 266, App
143	36	81.8	1089	14	US-10-176-913-266	Sequence 266, App	216	36	81.8	1089	14	US-10-184-616-266	Sequence 266, App
144	36	81.8	1089	14	US-10-180-552-266	Sequence 266, App	217	36	81.8	1089	14	US-10-184-617-266	Sequence 266, App
145	36	81.8	1089	14	US-10-180-557-266	Sequence 266, App	218	36	81.8	1089	14	US-10-184-622-266	Sequence 266, App
146	36	81.8	1089	14	US-10-173-700-266	Sequence 266, App	219	36	81.8	1089	14	US-10-184-628-266	Sequence 266, App
147	36	81.8	1089	14	US-10-174-572-266	Sequence 266, App	220	36	81.8	1089	14	US-10-184-629-266	Sequence 266, App
148	36	81.8	1089	14	US-10-174-579-266	Sequence 266, App	221	36	81.8	1089	14	US-10-184-630-266	Sequence 266, App
149	36	81.8	1089	14	US-10-174-582-266	Sequence 266, App	222	36	81.8	1089	14	US-10-184-631-266	Sequence 266, App
150	36	81.8	1089	14	US-10-174-588-266	Sequence 266, App	223	36	81.8	1089	14	US-10-184-633-266	Sequence 266, App
151	36	81.8	1089	14	US-10-175-739-266	Sequence 266, App	224	36	81.8	1089	14	US-10-184-636-266	Sequence 266, App
152	36	81.8	1089	14	US-10-175-740-266	Sequence 266, App	225	36	81.8	1089	14	US-10-184-640-266	Sequence 266, App
153	36	81.8	1089	14	US-10-175-743-266	Sequence 266, App	226	36	81.8	1089	14	US-10-184-650-266	Sequence 266, App
154	36	81.8	1089	14	US-10-176-488-266	Sequence 266, App	227	36	81.8	1089	14	US-10-184-651-266	Sequence 266, App
155	36	81.8	1089	14	US-10-176-492-266	Sequence 266, App	228	36	81.8	1089	14	US-10-187-588-266	Sequence 266, App
156	36	81.8	1089	14	US-10-176-747-266	Sequence 266, App	229	36	81.8	1089	14	US-10-187-597-266	Sequence 266, App
157	36	81.8	1089	14	US-10-176-750-266	Sequence 266, App	230	36	81.8	1089	14	US-10-187-598-266	Sequence 266, App
158	36	81.8	1089	14	US-10-176-985-266	Sequence 266, App	231	36	81.8	1089	14	US-10-187-600-266	Sequence 266, App
159	36	81.8	1089	14	US-10-176-987-266	Sequence 266, App	232	36	81.8	1089	14	US-10-187-601-266	Sequence 266, App
160	36	81.8	1089	14	US-10-176-992-266	Sequence 266, App	233	36	81.8	1089	14	US-10-187-602-266	Sequence 266, App
161	36	81.8	1089	14	US-10-176-993-266	Sequence 266, App	234	36	81.8	1089	14	US-10-187-603-266	Sequence 266, App

235	35	81.8	1089	14	US-10-187-741-266	Sequence 266, App	308	36	81.8	1089	14	US-10-194-432-266	Sequence 266, App
236	36	81.8	1089	14	US-10-187-743-266	Sequence 266, App	309	36	81.8	1089	14	US-10-194-483-266	Sequence 266, App
237	36	81.8	1089	14	US-10-187-746-266	Sequence 266, App	310	36	81.8	1089	14	US-10-195-883-266	Sequence 266, App
238	36	81.8	1089	14	US-10-187-747-266	Sequence 266, App	311	36	81.8	1089	14	US-10-195-889-266	Sequence 266, App
239	36	81.8	1089	14	US-10-187-751-266	Sequence 266, App	312	36	81.8	1089	14	US-10-196-749-266	Sequence 266, App
240	36	81.8	1089	14	US-10-187-753-266	Sequence 266, App	313	36	81.8	1089	14	US-10-196-750-266	Sequence 266, App
241	36	81.8	1089	14	US-10-187-754-266	Sequence 266, App	314	36	81.8	1089	14	US-10-197-699-266	Sequence 266, App
242	36	81.8	1089	14	US-10-187-757-266	Sequence 266, App	315	36	81.8	1089	14	US-10-197-700-266	Sequence 266, App
243	36	81.8	1089	14	US-10-187-884-266	Sequence 266, App	316	36	81.8	1089	14	US-10-197-703-266	Sequence 266, App
244	36	81.8	1089	14	US-10-188-767-266	Sequence 266, App	317	36	81.8	1089	14	US-10-197-708-266	Sequence 266, App
245	36	81.8	1089	14	US-10-188-769-266	Sequence 266, App	318	36	81.8	1089	14	US-10-198-764-266	Sequence 266, App
246	36	81.8	1089	14	US-10-188-770-266	Sequence 266, App	319	36	81.8	1089	14	US-10-198-765-266	Sequence 266, App
247	36	81.8	1089	14	US-10-188-773-266	Sequence 266, App	320	36	81.8	1089	14	US-10-198-768-266	Sequence 266, App
248	36	81.8	1089	14	US-10-188-781-266	Sequence 266, App	321	36	81.8	1089	14	US-10-198-769-266	Sequence 266, App
249	36	81.8	1089	14	US-10-194-361-266	Sequence 266, App	322	36	81.8	1089	14	US-10-199-303-266	Sequence 266, App
250	36	81.8	1089	14	US-10-194-362-266	Sequence 266, App	323	36	81.8	1089	14	US-10-199-306-266	Sequence 266, App
251	36	81.8	1089	14	US-10-195-893-266	Sequence 266, App	324	36	81.8	1089	14	US-10-199-310-266	Sequence 266, App
252	36	81.8	1089	14	US-10-195-901-266	Sequence 266, App	325	36	81.8	1089	14	US-10-199-311-266	Sequence 266, App
253	36	81.8	1089	14	US-10-195-902-266	Sequence 266, App	326	36	81.8	1089	14	US-10-199-314-266	Sequence 266, App
254	36	81.8	1089	14	US-10-196-743-266	Sequence 266, App	327	36	81.8	1089	14	US-10-199-317-266	Sequence 266, App
255	36	81.8	1089	14	US-10-196-760-266	Sequence 266, App	328	36	81.8	1089	14	US-10-199-665-266	Sequence 266, App
256	36	81.8	1089	14	US-10-173-708-266	Sequence 266, App	329	36	81.8	1089	14	US-10-199-666-266	Sequence 266, App
257	36	81.8	1089	14	US-10-176-479-266	Sequence 266, App	330	36	81.8	1089	14	US-10-199-669-266	Sequence 266, App
258	36	81.8	1089	14	US-10-176-748-266	Sequence 266, App	331	36	81.8	1089	14	US-10-201-534-266	Sequence 266, App
259	36	81.8	1089	14	US-10-176-916-266	Sequence 266, App	332	36	81.8	1089	14	US-10-201-770-266	Sequence 266, App
260	36	81.8	1089	14	US-10-179-507-266	Sequence 266, App	333	36	81.8	1089	14	US-10-201-855-266	Sequence 266, App
261	36	81.8	1089	14	US-10-179-516-266	Sequence 266, App	334	36	81.8	1089	14	US-10-201-856-266	Sequence 266, App
262	36	81.8	1089	14	US-10-184-626-266	Sequence 266, App	335	36	81.8	1089	14	US-10-202-933-266	Sequence 266, App
263	36	81.8	1089	14	US-10-184-627-266	Sequence 266, App	336	36	81.8	1089	14	US-10-202-934-266	Sequence 266, App
264	36	81.8	1089	14	US-10-184-652-266	Sequence 266, App	337	36	81.8	1089	14	US-10-202-935-266	Sequence 266, App
265	36	81.8	1089	14	US-10-184-654-266	Sequence 266, App	338	36	81.8	1089	14	US-10-202-936-266	Sequence 266, App
266	36	81.8	1089	14	US-10-183-008-266	Sequence 266, App	339	36	81.8	1089	14	US-10-202-937-266	Sequence 266, App
267	36	81.8	1089	14	US-10-183-009-266	Sequence 266, App	340	36	81.8	1089	14	US-10-202-938-266	Sequence 266, App
268	36	81.8	1089	14	US-10-183-017-266	Sequence 266, App	341	36	81.8	1089	14	US-10-202-939-266	Sequence 266, App
269	36	81.8	1089	14	US-10-183-019-266	Sequence 266, App	342	36	81.8	1089	14	US-10-205-504-266	Sequence 266, App
270	36	81.8	1089	14	US-10-184-618-266	Sequence 266, App	343	36	81.8	1089	14	US-10-205-509-266	Sequence 266, App
271	36	81.8	1089	14	US-10-184-625-266	Sequence 266, App	344	36	81.8	1089	14	US-10-205-895-266	Sequence 266, App
272	36	81.8	1089	14	US-10-184-626-266	Sequence 266, App	345	36	81.8	1089	14	US-10-205-899-266	Sequence 266, App
273	36	81.8	1089	14	US-10-184-627-266	Sequence 266, App	346	36	81.8	1089	14	US-10-205-900-266	Sequence 266, App
274	36	81.8	1089	14	US-10-184-645-266	Sequence 266, App	347	36	81.8	1089	14	US-10-205-909-266	Sequence 266, App
275	36	81.8	1089	14	US-10-184-654-266	Sequence 266, App	348	36	81.8	1089	14	US-10-195-890-266	Sequence 266, App
276	36	81.8	1089	14	US-10-184-655-266	Sequence 266, App	349	36	81.8	1089	14	US-10-006-818A-102	Sequence 102, App
277	36	81.8	1089	14	US-10-188-774-266	Sequence 266, App	350	36	81.8	1089	14	US-10-183-002-266	Sequence 266, App
278	36	81.8	1089	14	US-10-188-775-266	Sequence 266, App	351	36	81.8	1089	14	US-10-184-621-266	Sequence 266, App
279	36	81.8	1089	14	US-10-194-462-266	Sequence 266, App	352	36	81.8	1089	14	US-10-184-632-266	Sequence 266, App
280	36	81.8	1089	14	US-10-196-745-266	Sequence 266, App	353	36	81.8	1089	14	US-10-187-752-266	Sequence 266, App
281	36	81.8	1089	14	US-10-196-762-266	Sequence 266, App	354	36	81.8	1089	14	US-10-187-887-266	Sequence 266, App
282	36	81.8	1089	14	US-10-197-695-266	Sequence 266, App	355	36	81.8	1089	14	US-10-194-461-266	Sequence 266, App
283	36	81.8	1089	14	US-10-195-894-266	Sequence 102, App	356	36	81.8	1089	14	US-10-196-751-266	Sequence 266, App
284	36	81.8	1089	14	US-10-176-484-266	Sequence 266, App	357	36	81.8	1089	14	US-10-197-694-266	Sequence 266, App
285	36	81.8	1089	14	US-10-176-753-266	Sequence 266, App	358	36	81.8	1089	14	US-10-197-697-266	Sequence 266, App
286	36	81.8	1089	14	US-10-176-917-266	Sequence 266, App	359	36	81.8	1089	14	US-10-197-707-266	Sequence 266, App
287	36	81.8	1089	14	US-10-176-962-266	Sequence 266, App	360	36	81.8	1089	14	US-10-199-303-266	Sequence 266, App
288	36	81.8	1089	14	US-10-179-506-266	Sequence 266, App	361	36	81.8	1089	14	US-10-199-306-266	Sequence 266, App
289	36	81.8	1089	14	US-10-179-513-266	Sequence 266, App	362	36	81.8	1089	14	US-10-199-310-266	Sequence 266, App
290	36	81.8	1089	14	US-10-179-513-266	Sequence 266, App	363	36	81.8	1089	14	US-10-199-311-266	Sequence 266, App
291	36	81.8	1089	14	US-10-179-514-266	Sequence 266, App	364	36	81.8	1089	14	US-10-199-462-266	Sequence 266, App
292	36	81.8	1089	14	US-10-179-522-266	Sequence 266, App	365	36	81.8	1089	14	US-10-201-328-266	Sequence 266, App
293	36	81.8	1089	14	US-10-180-556-266	Sequence 266, App	366	36	81.8	1089	14	US-10-201-529-266	Sequence 266, App
294	36	81.8	1089	14	US-10-180-556-266	Sequence 266, App	367	36	81.8	1089	14	US-10-201-529-266	Sequence 266, App
295	36	81.8	1089	14	US-10-183-015-266	Sequence 266, App	368	36	81.8	1089	14	US-10-201-529-266	Sequence 266, App
296	36	81.8	1089	14	US-10-184-615-266	Sequence 266, App	369	36	81.8	1089	14	US-10-201-529-266	Sequence 266, App
297	36	81.8	1089	14	US-10-184-620-266	Sequence 266, App	370	36	81.8	1089	14	US-10-201-529-266	Sequence 266, App
298	36	81.8	1089	14	US-10-184-620-266	Sequence 266, App	371	36	81.8	1089	14	US-10-202-408-266	Sequence 266, App
299	36	81.8	1089	14	US-10-184-643-266	Sequence 266, App	372	36	81.8	1089	14	US-10-202-409-266	Sequence 266, App
300	36	81.8	1089	14	US-10-184-656-266	Sequence 266, App	373	36	81.8	1089	14	US-10-202-411-266	Sequence 266, App
301	36	81.8	1089	14	US-10-192-010-266	Sequence 266, App	374	36	81.8	1089	14	US-10-202-412-266	Sequence 266, App
302	36	81.8	1089	14	US-10-205-908-266	Sequence 266, App	375	36	81.8	1089	14	US-10-202-417-266	Sequence 266, App
303	36	81.8	1089	14	US-10-187-559-266	Sequence 266, App	376	36	81.8	1089	14	US-10-205-502-266	Sequence 266, App
304	36	81.8	1089	14	US-10-187-559-266	Sequence 266, App	377	36	81.8	1089	14	US-10-205-511-266	Sequence 266, App
305	36	81.8	1089	14	US-10-187-750-266	Sequence 266, App	378	36	81.8	1089	14	US-10-205-902-266	Sequence 266, App
306	36	81.8	1089	14	US-10-188-780-266	Sequence 266, App	379	36	81.8	1089	14	US-10-205-907-266	Sequence 266, App
307	36	81.8	1089	14	US-10-192-015-266	Sequence 266, App	380	36	81.8	1089	14	US-10-194-455-266	Sequence 266, App

381	36	81.8	1089	14	US-10-196-758-266	Sequence 266, App	454	36	81.8	1089	14	US-10-205-894-266	Sequence 266, App
382	36	81.8	1089	14	US-10-198-770-266	Sequence 266, App	455	36	81.8	1089	14	US-10-205-896-266	Sequence 266, App
383	36	81.8	1089	14	US-10-199-308-266	Sequence 266, App	456	36	81.8	1089	14	US-10-205-898-266	Sequence 266, App
384	36	81.8	1089	14	US-10-200-617-266	Sequence 266, App	457	36	81.8	1089	14	US-10-205-901-266	Sequence 266, App
385	36	81.8	1089	14	US-10-205-893-266	Sequence 266, App	458	36	81.8	1089	14	US-10-205-903-266	Sequence 266, App
386	36	81.8	1089	14	US-10-205-897-266	Sequence 266, App	459	36	81.8	1089	14	US-10-206-909-266	Sequence 266, App
387	36	81.8	1089	14	US-10-196-754-266	Sequence 266, App	460	36	81.8	1089	14	US-10-206-910-266	Sequence 266, App
388	36	81.8	1089	14	US-10-174-571-266	Sequence 266, App	461	36	81.8	1089	14	US-10-206-911-266	Sequence 266, App
389	36	81.8	1089	14	US-10-176-746-266	Sequence 266, App	462	36	81.8	1089	14	US-10-206-912-266	Sequence 266, App
390	36	81.8	1089	14	US-10-176-923-266	Sequence 266, App	463	36	81.8	1089	14	US-10-206-913-266	Sequence 266, App
391	36	81.8	1089	14	US-10-183-011-266	Sequence 266, App	464	36	81.8	1089	14	US-10-206-914-266	Sequence 266, App
392	36	81.8	1089	14	US-10-184-633-266	Sequence 266, App	465	36	81.8	1089	14	US-10-206-920-266	Sequence 266, App
393	36	81.8	1089	14	US-10-184-639-266	Sequence 266, App	466	36	81.8	1089	14	US-10-206-921-266	Sequence 266, App
394	36	81.8	1089	14	US-10-187-742-266	Sequence 266, App	467	36	81.8	1089	14	US-10-206-923-266	Sequence 266, App
395	36	81.8	1089	14	US-10-187-748-266	Sequence 266, App	468	36	81.8	1089	14	US-10-206-925-266	Sequence 266, App
396	36	81.8	1089	14	US-10-188-766-266	Sequence 266, App	469	36	81.8	1089	14	US-10-206-926-266	Sequence 266, App
397	36	81.8	1089	14	US-10-188-771-266	Sequence 266, App	470	36	81.8	1089	14	US-10-206-927-266	Sequence 266, App
398	36	81.8	1089	14	US-10-192-006-266	Sequence 266, App	471	36	81.8	1089	14	US-10-207-916-266	Sequence 266, App
399	36	81.8	1089	14	US-10-192-008-266	Sequence 266, App	472	36	81.8	1089	14	US-10-207-917-266	Sequence 266, App
400	36	81.8	1089	14	US-10-192-009-266	Sequence 266, App	473	36	81.8	1089	14	US-10-207-918-266	Sequence 266, App
401	36	81.8	1089	14	US-10-192-012-266	Sequence 266, App	474	36	81.8	1089	14	US-10-207-919-266	Sequence 266, App
402	36	81.8	1089	14	US-10-192-014-266	Sequence 266, App	475	36	81.8	1089	14	US-10-207-920-266	Sequence 266, App
403	36	81.8	1089	14	US-10-192-016-266	Sequence 266, App	476	36	81.8	1089	14	US-10-207-925-266	Sequence 266, App
404	36	81.8	1089	14	US-10-194-362-266	Sequence 266, App	477	36	81.8	1089	14	US-10-208-021-266	Sequence 266, App
405	36	81.8	1089	14	US-10-194-364-266	Sequence 266, App	478	36	81.8	1089	14	US-10-208-022-266	Sequence 266, App
406	36	81.8	1089	14	US-10-194-395-266	Sequence 266, App	479	36	81.8	1089	14	US-10-208-023-266	Sequence 266, App
407	36	81.8	1089	14	US-10-194-424-266	Sequence 266, App	480	36	81.8	1089	14	US-10-208-026-266	Sequence 266, App
408	36	81.8	1089	14	US-10-194-458-266	Sequence 266, App	481	36	81.8	1089	14	US-10-208-029-266	Sequence 266, App
409	36	81.8	1089	14	US-10-194-459-266	Sequence 266, App	482	36	81.8	1089	14	US-10-208-030-266	Sequence 266, App
410	36	81.8	1089	14	US-10-194-488-266	Sequence 266, App	483	36	81.8	1089	14	US-10-015-399A-102	Sequence 102, App
411	36	81.8	1089	14	US-10-195-886-266	Sequence 266, App	484	36	81.8	1089	14	US-10-232-232-266	Sequence 266, App
412	36	81.8	1089	14	US-10-195-891-266	Sequence 266, App	485	36	81.8	1089	14	US-10-195-898-266	Sequence 266, App
413	36	81.8	1089	14	US-10-196-746-266	Sequence 266, App	486	36	81.8	1089	14	US-10-196-759-266	Sequence 266, App
414	36	81.8	1089	14	US-10-196-752-266	Sequence 266, App	487	36	81.8	1089	14	US-10-196-759A-102	Sequence 102, App
415	36	81.8	1089	14	US-10-196-753-266	Sequence 266, App	488	36	81.8	1089	14	US-10-173-693-266	Sequence 266, App
416	36	81.8	1089	14	US-10-196-761-266	Sequence 266, App	489	36	81.8	1089	14	US-10-174-578-266	Sequence 266, App
417	36	81.8	1089	14	US-10-197-692-266	Sequence 266, App	490	36	81.8	1089	14	US-10-175-741-266	Sequence 266, App
418	36	81.8	1089	14	US-10-197-693-266	Sequence 266, App	491	36	81.8	1089	14	US-10-175-750-266	Sequence 266, App
419	36	81.8	1089	14	US-10-197-696-266	Sequence 266, App	492	36	81.8	1089	14	US-10-176-986-266	Sequence 266, App
420	36	81.8	1089	14	US-10-197-698-266	Sequence 266, App	493	36	81.8	1089	14	US-10-184-641-266	Sequence 266, App
421	36	81.8	1089	14	US-10-197-703-266	Sequence 266, App	494	36	81.8	1089	14	US-10-187-888-266	Sequence 266, App
422	36	81.8	1089	14	US-10-197-711-266	Sequence 266, App	495	36	81.8	1089	14	US-10-194-360-266	Sequence 266, App
423	36	81.8	1089	14	US-10-198-757-266	Sequence 266, App	496	36	81.8	1089	14	US-10-194-365-266	Sequence 266, App
424	36	81.8	1089	14	US-10-198-761-266	Sequence 266, App	497	36	81.8	1089	14	US-10-195-895-266	Sequence 266, App
425	36	81.8	1089	14	US-10-198-762-266	Sequence 266, App	498	36	81.8	1089	14	US-10-199-302-266	Sequence 266, App
426	36	81.8	1089	14	US-10-198-763-266	Sequence 266, App	499	36	81.8	1089	14	US-10-201-323-266	Sequence 266, App
427	36	81.8	1089	14	US-10-198-767-266	Sequence 266, App	500	36	81.8	1089	14	US-10-205-510-266	Sequence 266, App
428	36	81.8	1089	14	US-10-199-301-266	Sequence 266, App	ALIGNMENTS						
429	36	81.8	1089	14	US-10-199-307-266	Sequence 266, App							
430	36	81.8	1089	14	US-10-199-312-266	Sequence 266, App							
431	36	81.8	1089	14	US-10-199-315-266	Sequence 266, App							
432	36	81.8	1089	14	US-10-199-316-266	Sequence 266, App							
433	36	81.8	1089	14	US-10-199-457-266	Sequence 266, App							
434	36	81.8	1089	14	US-10-199-459-266	Sequence 266, App							
435	36	81.8	1089	14	US-10-199-460-266	Sequence 266, App							
436	36	81.8	1089	14	US-10-199-461-266	Sequence 266, App							
437	36	81.8	1089	14	US-10-199-667-266	Sequence 266, App							
438	36	81.8	1089	14	US-10-199-673-266	Sequence 266, App							
439	36	81.8	1089	14	US-10-201-321-266	Sequence 266, App							
440	36	81.8	1089	14	US-10-201-322-266	Sequence 266, App							
441	36	81.8	1089	14	US-10-201-326-266	Sequence 266, App							
442	36	81.8	1089	14	US-10-201-532-266	Sequence 266, App							
443	36	81.8	1089	14	US-10-201-533-266	Sequence 266, App							
444	36	81.8	1089	14	US-10-201-535-266	Sequence 266, App							
445	36	81.8	1089	14	US-10-201-769-266	Sequence 266, App							
446	36	81.8	1089	14	US-10-201-771-266	Sequence 266, App							
447	36	81.8	1089	14	US-10-201-854-266	Sequence 266, App							
448	36	81.8	1089	14	US-10-202-410-266	Sequence 266, App							
449	36	81.8	1089	14	US-10-202-473-266	Sequence 266, App							
450	36	81.8	1089	14	US-10-202-474-266	Sequence 266, App							
451	36	81.8	1089	14	US-10-205-503-266	Sequence 266, App							
452	36	81.8	1089	14	US-10-205-512-266	Sequence 266, App							
453	36	81.8	1089	14	US-10-205-892-266	Sequence 266, App							

RESULT 1  
US-10-235-552-2  
Sequence 2, Application US/10235552  
Publication No. US20030027768A1  
GENERAL INFORMATION:  
APPLICANT: Mazar, Andrew P.  
TITLE OF INVENTION: ANTI-INVASIVE AND ANTI-ANGIOGENIC COMPOSITIONS AND METHODS  
FILE REFERENCE: 38369-183655  
CURRENT FILING DATE: 2002-09-06  
PRIOR APPLICATION NUMBER: 08/500,327  
PRIOR FILING DATE: 1997-07-25  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: corresponds to

OTHER INFORMATION: positions 136-143 of uPA  
FEATURE:  
NAME/KEY: MOD\_RES  
LOCATION: (1)\_RES  
OTHER INFORMATION: ACETYLATION  
FEATURE:  
NAME/KEY: MOD\_RES  
LOCATION: (8)\_RES  
OTHER INFORMATION: AMIDATION  
US-10-235-552-2

Query Match 100.0%; Score 44; DB 14; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1e+06;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPE 8  
Db 1 KPSSPPE 8

RESULT 2  
US-10-235-552-1  
Sequence 1, Application US/10235552  
Publication No. US20030027768A1  
GENERAL INFORMATION:  
APPLICANT: Mazar, Andrew P.  
TITLE OF INVENTION: ANTI-INVASIVE AND ANTI-ANGIOGENIC COMPOSITIONS AND METHODS  
FILE REFERENCE: 38369-183655  
CURRENT APPLICATION NUMBER: US/10/235,552  
CURRENT FILING DATE: 2002-09-06  
PRIOR APPLICATION NUMBER: 08/900,327  
PRIOR FILING DATE: 1997-07-25  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 10  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: corresponds to  
OTHER INFORMATION: positions 136-145 of the human protein urokinase  
OTHER INFORMATION: plasminogen activator (uPA)  
FEATURE:  
NAME/KEY: MOD\_RES  
LOCATION: (1)  
OTHER INFORMATION: ACETYLATION  
FEATURE:  
NAME/KEY: MOD\_RES  
LOCATION: (10)  
OTHER INFORMATION: AMIDATION  
US-10-235-552-1

Query Match 100.0%; Score 44; DB 14; Length 10;  
Best Local Similarity 100.0%; Pred. No. 3.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPE 8  
Db 1 KPSSPPE 8

RESULT 3  
US-09-880-503-9  
Sequence 9, Application US/09880503  
Patent No. US20020131964A1  
GENERAL INFORMATION:  
APPLICANT: CINES, Douglas B  
APPLICANT: HIGAZI, Abd Al-Roof  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING MUSCLE CELL AND  
FILE REFERENCE: 9596-331  
CURRENT APPLICATION NUMBER: US/09/880,503

CURRENT FILING DATE: 2001-06-13  
PRIOR APPLICATION NUMBER: US 60/212,847  
PRIOR FILING DATE: 2000-06-20  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 9  
LENGTH: 96  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-880-503-9

Query Match 100.0%; Score 44; DB 9; Length 96;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPE 8  
Db 89 KPSSPPE 96

RESULT 4  
US-09-880-503-8  
Sequence 8, Application US/09880503  
Patent No. US20020131964A1  
GENERAL INFORMATION:  
APPLICANT: CINES, Douglas B  
APPLICANT: HIGAZI, Abd Al-Roof  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING MUSCLE CELL AND  
FILE REFERENCE: 9596-331  
CURRENT APPLICATION NUMBER: US/09/880,503  
CURRENT FILING DATE: 2001-06-13  
PRIOR APPLICATION NUMBER: US 60/212,847  
PRIOR FILING DATE: 2000-06-20  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 8  
LENGTH: 143  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-880-503-8

Query Match 100.0%; Score 44; DB 9; Length 143;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPE 8  
Db 136 KPSSPPE 143

RESULT 5  
US-09-880-503-5  
Sequence 5, Application US/09880503  
Patent No. US20020131964A1  
GENERAL INFORMATION:  
APPLICANT: CINES, Douglas B  
APPLICANT: HIGAZI, Abd Al-Roof  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING MUSCLE CELL AND  
FILE REFERENCE: 9596-331  
CURRENT APPLICATION NUMBER: US/09/880,503  
CURRENT FILING DATE: 2001-06-13  
PRIOR APPLICATION NUMBER: US 60/212,847  
PRIOR FILING DATE: 2000-06-20  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 276  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-880-503-5

Query Match 100.0%; Score 44; DB 9; Length 276;  
 Best Local Similarity 100.0%; Pred. No. 87;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
 |||||  
 Db 1 KPSSPPE 8

# RESULT 6

US-09-880-503-7  
 ; Sequence 7, Application US/09880503  
 ; Patent No. US20020131964A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CINES, Douglas B  
 ; APPLICANT: HIGAZI, Abd Al-Roof  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING MUSCLE CELL AND  
 ; FILE REFERENCE: 9596-331  
 ; CURRENT APPLICATION NUMBER: US/09/880,503  
 ; CURRENT FILING DATE: 2001-06-13  
 ; PRIOR APPLICATION NUMBER: US 60/212,847  
 ; PRIOR FILING DATE: 2000-06-20  
 ; NUMBER OF SEQ ID NOS: 18  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 7  
 ; LENGTH: 323  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-880-503-7

Query Match 100.0%; Score 44; DB 9; Length 323;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
 |||||  
 Db 48 KPSSPPE 55

# RESULT 7

US-10-106-698-6266  
 ; Sequence 6266, Application US/10106698  
 ; Publication No. US20030109690A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ruben et al.  
 ; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide  
 ; FILE REFERENCE: PA005P1  
 ; CURRENT APPLICATION NUMBER: US/10/106,698  
 ; CURRENT FILING DATE: 2002-03-27  
 ; PRIOR APPLICATION NUMBER: PCT/US00/26524  
 ; PRIOR FILING DATE: 2000-09-28  
 ; PRIOR APPLICATION NUMBER: US 60/157,137  
 ; PRIOR FILING DATE: 1999-09-29  
 ; PRIOR APPLICATION NUMBER: US 60/163,280  
 ; PRIOR FILING DATE: 1999-11-03  
 ; NUMBER OF SEQ ID NOS: 8564  
 ; SOFTWARE: Patentin Ver. 3.0  
 ; SEQ ID NO 6266  
 ; LENGTH: 337  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-106-698-6266

Query Match 100.0%; Score 44; DB 14; Length 337;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
 |||||  
 Db 162 KPSSPPE 169

RESULT 8  
 US-10-264-049-2927  
 ; Sequence 2927, Application US/10264049  
 ; Publication No. US20040005579A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Birse et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PA133P1  
 ; CURRENT APPLICATION NUMBER: US/10/264,049  
 ; CURRENT FILING DATE: 2002-10-04  
 ; PRIOR APPLICATION NUMBER: PCT/US01/18569  
 ; PRIOR FILING DATE: 2001-06-07  
 ; PRIOR APPLICATION NUMBER: US 60/209,467  
 ; PRIOR FILING DATE: 2000-06-07  
 ; NUMBER OF SEQ ID NOS: 4360  
 ; SOFTWARE: Patentin Ver. 3.1  
 ; SEQ ID NO 2927  
 ; LENGTH: 337  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-264-049-2927

Query Match 100.0%; Score 44; DB 15; Length 337;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
 |||||  
 Db 162 KPSSPPE 169

# RESULT 9

US-09-880-503-3  
 ; Sequence 3, Application US/09880503  
 ; Patent No. US20020131964A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CINES, Douglas B  
 ; APPLICANT: HIGAZI, Abd Al-Roof  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING MUSCLE CELL AND  
 ; FILE REFERENCE: 9596-331  
 ; CURRENT APPLICATION NUMBER: US/09/880,503  
 ; CURRENT FILING DATE: 2001-06-13  
 ; PRIOR APPLICATION NUMBER: US 60/212,847  
 ; PRIOR FILING DATE: 2000-06-20  
 ; NUMBER OF SEQ ID NOS: 18  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 3  
 ; LENGTH: 411  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-880-503-3

Query Match 100.0%; Score 44; DB 9; Length 411;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
 |||||  
 Db 136 KPSSPPE 143

# RESULT 10

US-10-407-821-2  
 ; Sequence 2, Application US/10407821  
 ; Publication No. US20030219386A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: IDELL, STEVEN  
 ; TITLE OF INVENTION: INTRAPLEURAL SINGLE-CHAIN UROKINASE ALONE OR COMPLEXED  
 ; TITLE OF INVENTION: TO ITS SOLUBLE RECEPTOR PROTECTS AGAINST PLEURAL  
 ; FILE REFERENCE: UTSN:022US  
 ; CURRENT APPLICATION NUMBER: US/10/407,821



```

; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: 60/414,202
; PRIOR FILING DATE: 2002-09-27
; PRIOR APPLICATION NUMBER: 60/370,466
; PRIOR FILING DATE: 2002-04-05
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-407-821-2

```

```

Query Match
Best Local Similarity 100.0%; Score 44; DB 15; Length 411;
Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 KPSSPPEE 8
Db 136 KPSSPPEE 143

```

```

RESULT 11
US-10-372-966-6
; Sequence 6, Application US/10372966
; Publication No. US20040111765A1
; GENERAL INFORMATION:
; APPLICANT: OISHI, Karen K.
; APPLICANT: ZHOU, Da-Feng
; TITLE OF INVENTION: PRODUCTION OF UROKINASE IN PLANT-BASED EXPRESSION
; FILE REFERENCE: 048281/0107
; CURRENT APPLICATION NUMBER: US/10/372,966
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: US/09/344,376
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/091,911
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Human
; US-10-372-966-6

```

```

Query Match
Best Local Similarity 100.0%; Score 44; DB 16; Length 411;
Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 KPSSPPEE 8
Db 137 KPSSPPEE 144

```

```

RESULT 12
US-10-372-966-4
; Sequence 4, Application US/10372966
; Publication No. US20040111765A1
; GENERAL INFORMATION:
; APPLICANT: OISHI, Karen K.
; APPLICANT: ZHOU, Da-Feng
; TITLE OF INVENTION: PRODUCTION OF UROKINASE IN PLANT-BASED EXPRESSION
; FILE REFERENCE: 048281/0107
; CURRENT APPLICATION NUMBER: US/10/372,966
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: US/09/344,376
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/091,911
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.0

```

```

; SEQ ID NO 4
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Human
; US-10-372-966-4

```

```

Query Match
Best Local Similarity 100.0%; Score 44; DB 16; Length 430;
Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 KPSSPPEE 8
Db 156 KPSSPPEE 163

```

```

RESULT 13
US-09-264-468B-1
; Sequence 1, Application US/09264468B
; Patent No. US20020106775A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Jieyi
; APPLICANT: Niemeyer, Vicki L.
; APPLICANT: Henkin, Jack
; APPLICANT: Smith, Richard A.
; APPLICANT: Walter, Karl A.
; APPLICANT: Severin, Jean M.
; APPLICANT: Edalji, Rohinton
; APPLICANT: Johnson Jr., Robert W.
; APPLICANT: Holzman, Thomas F.
; TITLE OF INVENTION: HIGHLY CRYSTALLINE UROKINASE
; FILE REFERENCE: 6310.US.P1
; CURRENT APPLICATION NUMBER: US/09/264,468B
; CURRENT FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: US 09/036,361
; PRIOR FILING DATE: 1998-03-06
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1) ... (20)
; OTHER INFORMATION: Leader sequence
; NAME/KEY: VARIANT
; LOCATION: (279) ... (279)
; OTHER INFORMATION: Xaa = any amino acid
; NAME/KEY: VARIANT
; LOCATION: (302) ... (302)
; OTHER INFORMATION: Xaa = any amino acid
; US-09-264-468B-1

```

```

Query Match
Best Local Similarity 100.0%; Score 44; DB 9; Length 431;
Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 KPSSPPEE 8
Db 156 KPSSPPEE 163

```

```

RESULT 14
US-10-411-037-34
; Sequence 34, Application US/10411037
; Publication No. US20040043446A1
; GENERAL INFORMATION:
; APPLICANT: Neose Technologies, Inc.
; APPLICANT: Defrees, Shawn
; APPLICANT: Zopf, David
; APPLICANT: Bayer, Robert
; APPLICANT: Hakes, David
; APPLICANT: Chen, Xi

```

APPLICANT: Bove, Caryn  
 TITLE OF INVENTION: ALPHA GALACTOSIDASE A: REMODELING AND GLYCOCONJUGATION OF ALPHA  
 FILE REFERENCE: 040853-01-5082  
 CURRENT APPLICATION NUMBER: US/10/411,037  
 PRIOR FILING DATE: 2003-04-09  
 PRIOR APPLICATION NUMBER: US 60/348,523  
 PRIOR FILING DATE: 2001-10-10  
 PRIOR APPLICATION NUMBER: US 60/344,692  
 PRIOR FILING DATE: 2001-10-19  
 PRIOR APPLICATION NUMBER: US 60/387,292  
 PRIOR FILING DATE: 2002-06-07  
 PRIOR APPLICATION NUMBER: US 60/391,777  
 PRIOR FILING DATE: 2002-06-25  
 PRIOR APPLICATION NUMBER: US 60/396,594  
 PRIOR FILING DATE: 2002-07-17  
 PRIOR APPLICATION NUMBER: US 60/404,249  
 PRIOR FILING DATE: 2002-08-16  
 PRIOR APPLICATION NUMBER: US 60/407,527  
 PRIOR FILING DATE: 2002-08-28  
 NUMBER OF SEQ ID NOS: 75  
 SOFTWARE: PatentIn version 3.2  
 SEQ ID NO 34  
 LENGTH: 431  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-411-037-34

Query Match 100.0%; Score 44; DB 12; Length 431;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
 Db 156 KPSSPPE 163

RESULT 15  
 US-10-411-026-34  
 Sequence 34, Application US/10411026  
 Publication No. US20040063911A1  
 GENERAL INFORMATION:  
 APPLICANT: Neose Technologies, Inc.  
 APPLICANT: Defrees, Shawn  
 APPLICANT: Zopf, David  
 APPLICANT: Bayer, Robert  
 APPLICANT: Hakes, David  
 APPLICANT: Chen, Xi  
 TITLE OF INVENTION: PROTEIN REMODELING METHODS AND PROTEINS/PEPTIDES PRODUCED BY THE  
 FILE REFERENCE: 040853-01-5053  
 CURRENT APPLICATION NUMBER: US/10/411,026  
 CURRENT FILING DATE: 2003-04-09  
 PRIOR APPLICATION NUMBER: US 60/328,523  
 PRIOR FILING DATE: 2001-10-10  
 PRIOR APPLICATION NUMBER: US 60/344,692  
 PRIOR FILING DATE: 2001-10-19  
 PRIOR APPLICATION NUMBER: US 60/387,292  
 PRIOR FILING DATE: 2002-06-07  
 PRIOR APPLICATION NUMBER: US 60/391,777  
 PRIOR FILING DATE: 2002-06-25  
 PRIOR APPLICATION NUMBER: US 60/396,594  
 PRIOR FILING DATE: 2002-07-17  
 PRIOR APPLICATION NUMBER: US 60/404,249  
 PRIOR FILING DATE: 2002-08-16  
 PRIOR APPLICATION NUMBER: US 60/407,527  
 PRIOR FILING DATE: 2002-08-28  
 NUMBER OF SEQ ID NOS: 75  
 SOFTWARE: PatentIn version 3.2  
 SEQ ID NO 34  
 LENGTH: 431  
 TYPE: PRT  
 ORGANISM: Homo sapiens

US-10-411-026-34  
 Query Match 100.0%; Score 44; DB 12; Length 431;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
 Db 156 KPSSPPE 163

RESULT 16  
 US-10-282-174-562  
 Sequence 562, Application US/10282174  
 Publication No. US20030224380A1  
 GENERAL INFORMATION:  
 APPLICANT: Becker, Kenneth David  
 APPLICANT: Velicelebi, Gonul  
 APPLICANT: Elliot, Kathryn J.  
 APPLICANT: Wang, Xin  
 APPLICANT: Tanzi, Rudolph E.  
 APPLICANT: Bertram, Lars  
 APPLICANT: Saunders, Aleister J.  
 APPLICANT: Mullin, Kristina M.  
 APPLICANT: Sampson, Andrew Johnson  
 APPLICANT: Blacker, Deborah Lynne  
 TITLE OF INVENTION: GENES AND POLYMORPHISMS ON CHROMOSOME 10  
 TITLE OF INVENTION: ASSOCIATED WITH ALZHEIMER'S DISEASE AND OTHER  
 FILE REFERENCE: 37481-3308  
 CURRENT APPLICATION NUMBER: US/10/282,174  
 CURRENT FILING DATE: 2002-10-25  
 PRIOR APPLICATION NUMBER: US 60/339,525  
 PRIOR FILING DATE: 2001-10-25  
 PRIOR APPLICATION NUMBER: US 60/338,010  
 PRIOR FILING DATE: 2001-11-08  
 PRIOR APPLICATION NUMBER: US 60/336,929  
 PRIOR FILING DATE: 2001-11-08  
 PRIOR APPLICATION NUMBER: US 60/338,363  
 PRIOR FILING DATE: 2001-11-09  
 PRIOR APPLICATION NUMBER: US 60/337,052  
 PRIOR FILING DATE: 2001-12-04  
 PRIOR APPLICATION NUMBER: US 60/368,919  
 PRIOR FILING DATE: 2002-03-28  
 NUMBER OF SEQ ID NOS: 564  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 562  
 LENGTH: 431  
 TYPE: PRT  
 ORGANISM: Homo sapiens

FEATURE:  
 NAME/KEY: VARIANT  
 LOCATION: 15, 58, 141, 214, 231, 274, 366  
 OTHER INFORMATION: xaa = Any Amino Acid  
 US-10-282-174-562

Query Match 100.0%; Score 44; DB 12; Length 431;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
 Db 156 KPSSPPE 163

RESULT 17  
 US-10-076-421-2  
 Sequence 2, Application US/10076421  
 Publication No. US20020193304A1  
 GENERAL INFORMATION:  
 APPLICANT: WADA, MANABU  
 APPLICANT: WADA, NAOKO  
 TITLE OF INVENTION: ANTI-HIV AGENTS

```
FILE REFERENCE: HAYAK-9
CURRENT APPLICATION NUMBER: US/10/076,421
CURRENT FILING DATE: 2002-05-17
PRIOR APPLICATION NUMBER: JP 2001-42655
PRIOR FILING DATE: 2001-02-20
PRIOR APPLICATION NUMBER: JP 2001-184284
PRIOR FILING DATE: 2001-06-19
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 431
TYPE: PRT
ORGANISM: Homo sapiens
US-10-076-421-2

Query Match      100.0%; Score 44; DB 13; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KPSSPPEE 8
Db      156 KPSSPPEE 163

RESULT 18
US-10-171-311-184
Sequence 184, Application US/10171311
Publication No. US20030087270A1
GENERAL INFORMATION:
APPLICANT: Schlegel, Robert
APPLICANT: Chen, Yan
APPLICANT: Zhao, Xumei
APPLICANT: Monahan, John
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Glatz, Karen
APPLICANT: Ganavavrapu, Manjula
APPLICANT: Hoersch, Sebastian
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
TITLE OF INVENTION: OF CERVICAL CANCER
FILE REFERENCE: MRI-035
CURRENT APPLICATION NUMBER: US/10/171,311
CURRENT FILING DATE: 2002-06-12
PRIOR APPLICATION NUMBER: US 60/298,159
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/298,155
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/335,936
PRIOR FILING DATE: 2001-11-14
NUMBER OF SEQ ID NOS: 238
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 184
LENGTH: 431
TYPE: PRT
ORGANISM: Homo sapiens
US-10-171-311-184

Query Match      100.0%; Score 44; DB 14; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KPSSPPEE 8
Db      156 KPSSPPEE 163

RESULT 19
US-10-193-656-4
Sequence 4, Application US/10193656
Publication No. US20030096732A1
GENERAL INFORMATION:
APPLICANT: NY, Tor
APPLICANT: HOLMDAHL, Rikard
```

```
APPLICANT: Li, Jinao
TITLE OF INVENTION: NOVEL DRUG TARGETS FOR ARTHRITIS
FILE REFERENCE: 3810/13577-US3
CURRENT APPLICATION NUMBER: US/10/193,656
CURRENT FILING DATE: 2002-07-10
PRIOR APPLICATION NUMBER: US 60/304,461
PRIOR FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: US 60/304,490
PRIOR FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: US 60/305,182
PRIOR FILING DATE: 2001-07-13
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 431
TYPE: PRT
ORGANISM: Homo sapiens
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: GenBank / P00749
DATABASE ENTRY DATE: 1986-07-21
RELEVANT RESIDUES: (1)..(431)
US-10-193-656-4

Query Match      100.0%; Score 44; DB 14; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KPSSPPEE 8
Db      156 KPSSPPEE 163
```

```
RESULT 20
US-10-301-822-161
Sequence 161, Application US/10301822
Publication No. US20030148410A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: Berger, Allison
APPLICANT: Guillemette, Tracy L.
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Schlegel, Robert
APPLICANT: Monahan, John E.
APPLICANT: Thibodeau, Stephen N.
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
TITLE OF INVENTION: THERAPY OF COLON CANCER
FILE REFERENCE: MPM01-0292PRM
CURRENT APPLICATION NUMBER: US/10/301,822
CURRENT FILING DATE: 2002-11-21
PRIOR APPLICATION NUMBER: US 60/339,971
PRIOR FILING DATE: 2001-12-10
PRIOR APPLICATION NUMBER: US 60/361,978
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/381,988
PRIOR FILING DATE: 2002-05-20
NUMBER OF SEQ ID NOS: 228
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 161
LENGTH: 431
TYPE: PRT
ORGANISM: Homo Sapiens
US-10-301-822-161

Query Match      100.0%; Score 44; DB 14; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KPSSPPEE 8
Db      156 KPSSPPEE 163
```

RESULT 21  
US-10-247-671-149  
; Sequence 149, Application US/10247671  
; Publication No. US20030194721A1  
; GENERAL INFORMATION:  
; APPLICANT: Mikita, Thomas  
; APPLICANT: Shiftman, Dov  
; APPLICANT: Porter, Gordon, J.  
; APPLICANT: Kaser, Matthew R.  
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS  
; FILE REFERENCE: PA-0050 US  
; CURRENT APPLICATION NUMBER: US/10/247,671  
; PRIOR FILING DATE: 2002-09-18  
; PRIOR FILING DATE: 2001-09-19  
; NUMBER OF SEQ ID NOS: 186  
; SOFTWARE: PERL Program  
; SEQ ID NO 149  
; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030194721A1 1453334CD1  
US-10-247-671-149

Query Match 100.0%; Score 44; DB 14; Length 431;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8  
Db 156 KPSSPPEE 163

RESULT 22  
US-10-131-985-21  
; Sequence 21, Application US/10131985  
; Publication No. US20030199440A1  
; GENERAL INFORMATION:  
; APPLICANT: Dack, Kevin N  
; APPLICANT: Davies, Michael J  
; APPLICANT: Fish, Paul V  
; APPLICANT: Huggins, Jonathan P  
; APPLICANT: McIntosh, Fraser S  
; APPLICANT: Ocleston, Nicholas L  
; TITLE OF INVENTION: Composition  
; FILE REFERENCE: PCS 10351A  
; CURRENT APPLICATION NUMBER: US/10/131,985  
; PRIOR FILING DATE: 2002-04-25  
; PRIOR APPLICATION NUMBER: US/09/726,295  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: GB 9930768.8  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: Patentln Ver. 2.1  
; SEQ ID NO 21  
; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-131-985-21

Query Match 100.0%; Score 44; DB 14; Length 431;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8  
Db 156 KPSSPPEE 163

RESULT 23

US-10-295-027-414  
; Sequence 414, Application US/10295027  
; Publication No. US20030232350A1  
; GENERAL INFORMATION:  
; APPLICANT: Afari, Daniel  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Ginsberg, Wendy M.  
; APPLICANT: Gish, Kurt C.  
; APPLICANT: Glynn, Richard  
; APPLICANT: Hevezi, Peter A.  
; APPLICANT: Mack, David H.  
; APPLICANT: Murray, Richard  
; APPLICANT: Watson, Susan R.  
; APPLICANT: Eos Biotechnology, Inc.  
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and  
; FILE REFERENCE: 018501-012500US  
; CURRENT APPLICATION NUMBER: US/10/295,027  
; CURRENT FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: US 09/663,733  
; PRIOR FILING DATE: 2000-09-15  
; PRIOR APPLICATION NUMBER: US 60/350,666  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/335,394  
; PRIOR FILING DATE: 2001-11-15  
; PRIOR APPLICATION NUMBER: US 60/332,464  
; PRIOR FILING DATE: 2001-11-21  
; PRIOR APPLICATION NUMBER: US 60/334,393  
; PRIOR FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: US 60/340,376  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: US 60/347,211  
; PRIOR FILING DATE: 2002-01-08  
; PRIOR APPLICATION NUMBER: US 60/347,349  
; PRIOR FILING DATE: 2002-01-10  
; PRIOR APPLICATION NUMBER: US 60/355,250  
; PRIOR FILING DATE: 2002-02-08  
; PRIOR APPLICATION NUMBER: US 60/356,714  
; PRIOR FILING DATE: 2002-02-13  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 1386  
; SOFTWARE: Patentln Ver. 2.1  
; SEQ ID NO 414  
; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-295-027-414

Query Match 100.0%; Score 44; DB 15; Length 431;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8  
Db 156 KPSSPPEE 163

RESULT 24  
US-10-295-027-1275  
; Sequence 1275, Application US/10295027  
; Publication No. US20030232350A1  
; GENERAL INFORMATION:  
; APPLICANT: Afari, Daniel  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Ginsberg, Wendy M.  
; APPLICANT: Gish, Kurt C.  
; APPLICANT: Glynn, Richard  
; APPLICANT: Hevezi, Peter A.  
; APPLICANT: Mack, David H.  
; APPLICANT: Murray, Richard  
; APPLICANT: Watson, Susan R.  
; APPLICANT: Eos Biotechnology, Inc.  
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and

TITLE OF INVENTION: Methods of Screening for Modulators of Cancer  
FILE REFERENCE: 018501-012500US  
CURRENT APPLICATION NUMBER: US/10/295,027  
CURRENT FILING DATE: 2002-11-13  
PRIOR APPLICATION NUMBER: US 09/663,733  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: US 60/350,666  
PRIOR FILING DATE: 2001-11-13  
PRIOR APPLICATION NUMBER: US 60/335,394  
PRIOR FILING DATE: 2001-11-15  
PRIOR APPLICATION NUMBER: US 60/332,464  
PRIOR FILING DATE: 2001-11-21  
PRIOR APPLICATION NUMBER: US 60/334,393  
PRIOR FILING DATE: 2001-11-29  
PRIOR APPLICATION NUMBER: US 60/340,376  
PRIOR FILING DATE: 2001-12-14  
PRIOR APPLICATION NUMBER: US 60/347,211  
PRIOR FILING DATE: 2002-01-08  
PRIOR APPLICATION NUMBER: US 60/347,349  
PRIOR FILING DATE: 2002-01-10  
PRIOR APPLICATION NUMBER: US 60/355,250  
PRIOR FILING DATE: 2002-02-08  
PRIOR APPLICATION NUMBER: US 60/356,714  
PRIOR FILING DATE: 2002-02-13  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 1386  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1275  
LENGTH: 431  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-295-027-1275

Query Match 100.0%; Score 44; DB 15; Length 431;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8  
Db 156 KPSSPPEE 163

RESULT 25  
US-10-410-962-34  
Sequence 34, Application US/10410962  
Publication No. US20040077836A1  
GENERAL INFORMATION:  
APPLICANT: Neose Technologies, Inc.  
APPLICANT: Defrees, Shawn  
APPLICANT: Zopf, David  
APPLICANT: Bayer, Robert  
APPLICANT: Hakes, David  
APPLICANT: Chen, Xi  
TITLE OF INVENTION: GRANULOCYTE COLONY STIMULATING FACTOR: REMODELING AND  
FILE REFERENCE: 040853-01-5054  
CURRENT APPLICATION NUMBER: US/10/410,962  
CURRENT FILING DATE: 2003-04-09  
PRIOR APPLICATION NUMBER: US 60/328,523  
PRIOR FILING DATE: 2001-10-10  
PRIOR APPLICATION NUMBER: US 60/344,692  
PRIOR FILING DATE: 2001-10-19  
PRIOR APPLICATION NUMBER: US 60/387,292  
PRIOR FILING DATE: 2002-06-07  
PRIOR APPLICATION NUMBER: US 60/391,777  
PRIOR FILING DATE: 2002-06-25  
PRIOR APPLICATION NUMBER: US 60/396,594  
PRIOR FILING DATE: 2002-07-17  
PRIOR APPLICATION NUMBER: US 60/404,249  
PRIOR FILING DATE: 2002-08-16  
PRIOR APPLICATION NUMBER: US 60/407,527  
PRIOR FILING DATE: 2002-08-28

NUMBER OF SEQ ID NOS: 75  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 34  
LENGTH: 431  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-410-962-34

Query Match 100.0%; Score 44; DB 16; Length 431;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8  
Db 156 KPSSPPEE 163

RESULT 26  
US-10-411-049-34  
Sequence 34, Application US/10411049  
Publication No. US20040082026A1  
GENERAL INFORMATION:  
APPLICANT: Neose Technologies, Inc.  
APPLICANT: Defrees, Shawn  
APPLICANT: Zopf, David  
APPLICANT: Bayer, Robert  
APPLICANT: Hakes, David  
APPLICANT: Chen, Xi  
TITLE OF INVENTION: INTERFERON ALPHA: REMODELING AND GLYCOCONJUGATION OF INTERFERON  
FILE REFERENCE: 040853-01-5055  
CURRENT APPLICATION NUMBER: US/10/411,049  
CURRENT FILING DATE: 2003-04-09  
PRIOR APPLICATION NUMBER: US 60/328,523  
PRIOR FILING DATE: 2001-10-10  
PRIOR APPLICATION NUMBER: US 60/344,692  
PRIOR FILING DATE: 2001-10-19  
PRIOR APPLICATION NUMBER: US 60/387,292  
PRIOR FILING DATE: 2002-06-07  
PRIOR APPLICATION NUMBER: US 60/391,777  
PRIOR FILING DATE: 2002-06-25  
PRIOR APPLICATION NUMBER: US 60/396,594  
PRIOR FILING DATE: 2002-07-17  
PRIOR APPLICATION NUMBER: US 60/404,249  
PRIOR FILING DATE: 2002-08-16  
PRIOR APPLICATION NUMBER: US 60/407,527  
PRIOR FILING DATE: 2002-08-28  
NUMBER OF SEQ ID NOS: 75  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 34  
LENGTH: 431  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-411-049-34

Query Match 100.0%; Score 44; DB 16; Length 431;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8  
Db 156 KPSSPPEE 163

RESULT 27  
US-10-372-966-5  
Sequence 5, Application US/10372966  
Publication No. US20040111765A1  
GENERAL INFORMATION:  
APPLICANT: OISHI, Karen K.  
APPLICANT: ZHOU, Da-Feng  
TITLE OF INVENTION: PRODUCTION OF UROKINASE IN PLANT-BASED EXPRESSION

```

; TITLE OF INVENTION: SYSTEMS
; FILE REFERENCE: 048281/0107
; CURRENT APPLICATION NUMBER: US/10/372,966
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: US/09/344,376
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/091,911
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Human and Potato
; US-10-372-966-5

Query Match      100.0%; Score 44; DB 16; Length 433;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KPSSPPE 8
Db      158 KPSSPPE 165

RESULT 28
US-10-087-192-594
; Sequence 594, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 594
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-087-192-594

Query Match      100.0%; Score 44; DB 12; Length 437;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KPSSPPE 8
Db      162 KPSSPPE 169

RESULT 29
US-10-360-101-266
; Sequence 266, Application US/10360101
; Publication No. US20040009550A1
; GENERAL INFORMATION:
; APPLICANT: Moill, Gert N.
; APPLICANT: Leenhouts, Cornelis J.
; TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic way
; FILE REFERENCE: 2183-5673
; CURRENT APPLICATION NUMBER: US/10/360,101
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: EP 02077060.8
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 309
; SOFTWARE: Patentin version 3.1
```

```

; SEQ ID NO 266
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: sequence of urokinase
; US-10-360-101-266

Query Match      100.0%; Score 44; DB 15; Length 445;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KPSSPPE 8
Db      160 KPSSPPE 167

RESULT 30
US-10-235-552-8
; Sequence 8, Application US/10235552
; Publication No. US20030027768A1
; GENERAL INFORMATION:
; APPLICANT: Mazar, Andrew P.
; APPLICANT: Jones, Terence L.
; TITLE OF INVENTION: ANTI-INVASIVE AND ANTI-ANGIOGENIC COMPOSITIONS AND METHODS
; FILE REFERENCE: 38369-183655
; CURRENT APPLICATION NUMBER: US/10/235,552
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 08/900,327
; PRIOR FILING DATE: 1997-07-25
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: substitution,
; OTHER INFORMATION: deletion or addition variants of SEQ ID NO:2
; US-10-235-552-8

Query Match      93.2%; Score 41; DB 14; Length 8;
Best Local Similarity 87.5%; Pred. No. 1e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 KPSSPPE 8
Db      1 RPSSPPE 8

RESULT 31
US-10-235-552-4
; Sequence 4, Application US/10235552
; Publication No. US20030027768A1
; GENERAL INFORMATION:
; APPLICANT: Mazar, Andrew P.
; APPLICANT: Jones, Terence L.
; TITLE OF INVENTION: ANTI-INVASIVE AND ANTI-ANGIOGENIC COMPOSITIONS AND METHODS
; FILE REFERENCE: 38369-183655
; CURRENT APPLICATION NUMBER: US/10/235,552
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 08/900,327
; PRIOR FILING DATE: 1997-07-25
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: substitution,
; OTHER INFORMATION: deletion or addition variants of SEQ ID NO:2
; FEATURE:
```

NAME/KEY: MOD\_RES  
LOCATION: (1)  
OTHER INFORMATION: ACETYLATION  
FEATURE:  
NAME/KEY: MOD\_RES  
LOCATION: (7)  
OTHER INFORMATION: AMIDATION  
US-10-235-552-4

Query Match 88.6%; Score 39; DB 14; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PSSPPE 8  
Db 1 PSSPPE 7

RESULT 32  
US-10-235-552-5  
Sequence 5, Application US/10235552  
Publication No. US20030027768A1  
GENERAL INFORMATION:  
APPLICANT: Mazar, Andrew P.  
APPLICANT: Jones, Terence L.  
TITLE OF INVENTION: ANTI-INVASIVE AND ANTI-ANGIOGENIC COMPOSITIONS AND METHODS  
FILE REFERENCE: 38369-183655  
CURRENT APPLICATION NUMBER: US/10/235,552  
CURRENT FILING DATE: 2002-09-06  
PRIOR APPLICATION NUMBER: 08/900,327  
PRIOR FILING DATE: 1997-07-25  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 7  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: substitution,  
OTHER INFORMATION: deletion or addition variants of SEQ ID NO:2  
FEATURE:  
NAME/KEY: MOD\_RES  
LOCATION: (1)  
OTHER INFORMATION: ACETYLATION  
FEATURE:  
NAME/KEY: MOD\_RES  
LOCATION: (7)  
OTHER INFORMATION: AMIDATION  
US-10-235-552-5

Query Match 88.6%; Score 39; DB 14; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSPPPE 7  
Db 1 KPSPPPE 7

RESULT 33  
US-10-235-552-6  
Sequence 6, Application US/10235552  
Publication No. US20030027768A1  
GENERAL INFORMATION:  
APPLICANT: Mazar, Andrew P.  
APPLICANT: Jones, Terence L.  
TITLE OF INVENTION: ANTI-INVASIVE AND ANTI-ANGIOGENIC COMPOSITIONS AND METHODS  
FILE REFERENCE: 38369-183655  
CURRENT APPLICATION NUMBER: US/10/235,552  
CURRENT FILING DATE: 2002-09-06  
PRIOR APPLICATION NUMBER: 08/900,327  
PRIOR FILING DATE: 1997-07-25  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 6  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: substitution,  
OTHER INFORMATION: deletion or addition variants of SEQ ID NO:2  
FEATURE:  
NAME/KEY: MOD\_RES  
LOCATION: (1)  
OTHER INFORMATION: ACETYLATION  
FEATURE:  
NAME/KEY: MOD\_RES  
LOCATION: (7)  
OTHER INFORMATION: AMIDATION  
US-10-235-552-6

Query Match 88.6%; Score 39; DB 14; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSPPPE 7  
Db 1 KPSPPPE 7

RESULT 34  
US-10-235-552-7  
Sequence 7, Application US/10235552  
Publication No. US20030027768A1  
GENERAL INFORMATION:  
APPLICANT: Mazar, Andrew P.  
APPLICANT: Jones, Terence L.  
TITLE OF INVENTION: ANTI-INVASIVE AND ANTI-ANGIOGENIC COMPOSITIONS AND METHODS  
FILE REFERENCE: 38369-183655  
CURRENT APPLICATION NUMBER: US/10/235,552  
CURRENT FILING DATE: 2002-09-06  
PRIOR APPLICATION NUMBER: 08/900,327  
PRIOR FILING DATE: 1997-07-25  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 7  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: substitution,  
OTHER INFORMATION: deletion or addition variants of SEQ ID NO:2  
FEATURE:  
NAME/KEY: MOD\_RES  
LOCATION: (1)  
OTHER INFORMATION: ACETYLATION  
FEATURE:  
NAME/KEY: MOD\_RES  
LOCATION: (8)  
OTHER INFORMATION: AMIDATION  
US-10-235-552-7

Query Match 86.4%; Score 38; DB 14; Length 8;  
Best Local Similarity 75.0%; Pred. No. 1e+06;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSPPPE 8  
Db 1 KPSPPPD 8

RESULT 35  
US-09-864-761-37883  
Sequence 37883, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Mazar, Andrew P.  
APPLICANT: Jones, Terence L.  
TITLE OF INVENTION: ANTI-INVASIVE AND ANTI-ANGIOGENIC COMPOSITIONS AND METHODS  
FILE REFERENCE: 38369-183655  
CURRENT APPLICATION NUMBER: US/10/235,552  
CURRENT FILING DATE: 2002-09-06  
PRIOR APPLICATION NUMBER: 08/900,327  
PRIOR FILING DATE: 1997-07-25  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 8  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: substitution,  
OTHER INFORMATION: deletion or addition variants of SEQ ID NO:2  
FEATURE:  
NAME/KEY: MOD\_RES  
LOCATION: (1)  
OTHER INFORMATION: ACETYLATION  
FEATURE:  
NAME/KEY: MOD\_RES  
LOCATION: (8)  
OTHER INFORMATION: AMIDATION  
US-10-235-552-8

SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 6  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: substitution,  
OTHER INFORMATION: deletion or addition variants of SEQ ID NO:2  
FEATURE:  
NAME/KEY: MOD\_RES  
LOCATION: (1)  
OTHER INFORMATION: ACETYLATION  
FEATURE:  
NAME/KEY: MOD\_RES  
LOCATION: (8)  
OTHER INFORMATION: AMIDATION  
US-10-235-552-6

Query Match 86.4%; Score 38; DB 14; Length 8;  
Best Local Similarity 75.0%; Pred. No. 1e+06;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSPPPE 8  
Db 1 KPTTPEE 8

RESULT 34  
US-10-235-552-7  
Sequence 7, Application US/10235552  
Publication No. US20030027768A1  
GENERAL INFORMATION:  
APPLICANT: Mazar, Andrew P.  
APPLICANT: Jones, Terence L.  
TITLE OF INVENTION: ANTI-INVASIVE AND ANTI-ANGIOGENIC COMPOSITIONS AND METHODS  
FILE REFERENCE: 38369-183655  
CURRENT APPLICATION NUMBER: US/10/235,552  
CURRENT FILING DATE: 2002-09-06  
PRIOR APPLICATION NUMBER: 08/900,327  
PRIOR FILING DATE: 1997-07-25  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 7  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: substitution,  
OTHER INFORMATION: deletion or addition variants of SEQ ID NO:2  
FEATURE:  
NAME/KEY: MOD\_RES  
LOCATION: (1)  
OTHER INFORMATION: ACETYLATION  
FEATURE:  
NAME/KEY: MOD\_RES  
LOCATION: (8)  
OTHER INFORMATION: AMIDATION  
US-10-235-552-7

Query Match 86.4%; Score 38; DB 14; Length 8;  
Best Local Similarity 75.0%; Pred. No. 1e+06;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSPPPE 8  
Db 1 KPSPPPD 8

RESULT 35  
US-09-864-761-37883  
Sequence 37883, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Mazar, Andrew P.  
APPLICANT: Jones, Terence L.  
TITLE OF INVENTION: ANTI-INVASIVE AND ANTI-ANGIOGENIC COMPOSITIONS AND METHODS  
FILE REFERENCE: 38369-183655  
CURRENT APPLICATION NUMBER: US/10/235,552  
CURRENT FILING DATE: 2002-09-06  
PRIOR APPLICATION NUMBER: 08/900,327  
PRIOR FILING DATE: 1997-07-25  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 8  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: substitution,  
OTHER INFORMATION: deletion or addition variants of SEQ ID NO:2  
FEATURE:  
NAME/KEY: MOD\_RES  
LOCATION: (1)  
OTHER INFORMATION: ACETYLATION  
FEATURE:  
NAME/KEY: MOD\_RES  
LOCATION: (8)  
OTHER INFORMATION: AMIDATION  
US-10-235-552-8

APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263,6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 37883  
LENGTH: 166  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC011232.3  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.6  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.7  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5  
OTHER INFORMATION: EST HUMAN HIT: AM444637.1, EVALUATE 1.00e-76  
OTHER INFORMATION: SWISSPROT HIT: Q92HB3, EVALUATE 2.80e+00  
US-09-864-761-37883

Query Match 86.4%; Score 38; DB 9; Length 166;  
Best Local Similarity 87.5%; Pred. No. 3.9e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 KPSSPPE 8  
|||||

Db 84 KPSSPPE 91

RESULT 36

US-09-864-761-46504

Sequence 46504, Application US/09864761  
Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

FILE REFERENCE: Aeomica-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263,6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/774,203

PRIOR FILING DATE: 2001-01-29

NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

SEQ ID NO 46504

LENGTH: 166

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:  
OTHER INFORMATION: MAP TO AL121653.1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.7  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 4.2  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.92  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.9  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6  
OTHER INFORMATION: EST HUMAN HIT: AM444637.1, EVALUATE 1.00e-76  
OTHER INFORMATION: SWISSPROT HIT: Q92HB3, EVALUATE 2.80e+00  
US-09-864-761-46504



Query Match 86.4%; Score 38; DB 9; Length 166;  
Best Local Similarity 87.5%; Pred. No. 3.9e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8  
Db 84 KPSSPPEE 91

RESULT 37  
US-10-108-260A-3871

Sequence 3871, Application US/10108260A  
Publication No. US20040005560A1  
GENERAL INFORMATION:  
APPLICANT: HELIX RESEARCH INSTITUTE  
TITLE OF INVENTION: No. US20040005560A1el full length cDNA  
FILE REFERENCE: H1-A0106  
CURRENT APPLICATION NUMBER: US/10/108,260A  
CURRENT FILING DATE: 2002-03-27  
NUMBER OF SEQ ID NOS: 5458  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3871  
LENGTH: 264  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-108-260A-3871

Query Match 86.4%; Score 38; DB 15; Length 264;  
Best Local Similarity 87.5%; Pred. No. 6e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8  
Db 182 KPSSPPEE 189

RESULT 38  
US-10-176-306-23

Sequence 23, Application US/10176306  
Publication No. US20030130485A1  
GENERAL INFORMATION:  
APPLICANT: Meyers, Rachel A.  
APPLICANT: Glucksmann, Maria Alexandra  
APPLICANT: Bandaru, Rajasekhar  
APPLICANT: Kapeller-Libermann, Rosana  
TITLE OF INVENTION: NOVEL HUMAN GENES AND METHODS OF USE THEREOF  
FILE REFERENCE: 10448-195001  
CURRENT APPLICATION NUMBER: US/10/176,306  
CURRENT FILING DATE: 2002-06-20  
PRIOR APPLICATION NUMBER: 10/001,137  
PRIOR FILING DATE: 2001-11-14  
PRIOR APPLICATION NUMBER: PCT/US01/45291  
PRIOR FILING DATE: 2001-11-14  
PRIOR APPLICATION NUMBER: 60/248,362  
PRIOR FILING DATE: 2000-11-14  
PRIOR APPLICATION NUMBER: 60/248,331  
PRIOR FILING DATE: 2000-11-14  
PRIOR APPLICATION NUMBER: 60/248,365  
PRIOR FILING DATE: 2000-11-14  
PRIOR APPLICATION NUMBER: 60/250,077  
PRIOR FILING DATE: 2000-11-30  
PRIOR APPLICATION NUMBER: 60/250,327  
PRIOR FILING DATE: 2000-11-30  
PRIOR APPLICATION NUMBER: 60/250,176  
PRIOR FILING DATE: 2000-11-30  
PRIOR APPLICATION NUMBER: 10/023,617  
PRIOR FILING DATE: 2001-12-18  
PRIOR APPLICATION NUMBER: PCT/US01/49416  
PRIOR FILING DATE: 2001-12-18  
PRIOR APPLICATION NUMBER: 60/256,249  
PRIOR FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: 60/256,405

PRIOR FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: 10/083,248  
PRIOR FILING DATE: 2001-10-22  
PRIOR APPLICATION NUMBER: PCT/US01/46717  
PRIOR FILING DATE: 2001-10-22  
PRIOR APPLICATION NUMBER: 60/242,324  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/242,518  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/241,989  
PRIOR FILING DATE: 2000-10-20  
NUMBER OF SEQ ID NOS: 86  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 23  
LENGTH: 461  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-176-306-23

Qy 1 KPSSPPEE 8  
Db 379 KPSSPPEE 386

RESULT 39  
US-10-094-749-2145

Sequence 2145, Application US/10094749  
Publication No. US20030219741A1  
GENERAL INFORMATION:  
APPLICANT: ISOGAI, TAKAO  
APPLICANT: SUGIYAMA, TOMOYASU  
APPLICANT: OTSUKI, TETSUJI  
APPLICANT: WAKAMATSU, AI  
APPLICANT: SATO, HIROYUKI  
APPLICANT: ISHII, SHIZUKO  
APPLICANT: YAMAMOTO, JUN-ICHI  
APPLICANT: ISONO, YUUKO  
APPLICANT: HIO, YURI  
APPLICANT: OTSUKA, KAORU  
APPLICANT: NAGAI, KEIICHI  
APPLICANT: IRIE, RYOTARO  
APPLICANT: TAMECHIKA, ICHIRO  
APPLICANT: SEKI, NAOHIKO  
APPLICANT: YOSHIKAWA, TSUTOMU  
APPLICANT: OTSUKA, MOTOMYUKI  
APPLICANT: NAGAHARI, KENJI  
APPLICANT: MASUHO, YASUHIKO  
TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA  
FILE REFERENCE: 084335/0160  
CURRENT APPLICATION NUMBER: US/10/094,749  
CURRENT FILING DATE: 2002-03-12  
PRIOR APPLICATION NUMBER: 60/350,435  
PRIOR FILING DATE: 2002-01-24  
PRIOR APPLICATION NUMBER: JP 2001-328381  
PRIOR FILING DATE: 2001-09-14  
NUMBER OF SEQ ID NOS: 3381  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2145  
LENGTH: 461  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-094-749-2145

Query Match 86.4%; Score 38; DB 15; Length 461;  
Best Local Similarity 87.5%; Pred. No. 1e+03;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8  
Db 182 KPSSPPEE 189

Db 379 KPSSPPE 386

RESULT 40

US-10-295-027-205  
 ; Sequence 205, Application US/10295027  
 ; Publication No. US2003023350A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Afar, Daniel  
 ; APPLICANT: Aziz, Natasha  
 ; APPLICANT: Ginsberg, Wendy M.  
 ; APPLICANT: Gish, Kurt C.  
 ; APPLICANT: Glynn, Richard  
 ; APPLICANT: Hevezi, Peter A.  
 ; APPLICANT: Mack, David H.  
 ; APPLICANT: Murray, Richard  
 ; APPLICANT: Watson, Susan R.  
 ; APPLICANT: Eos Biotechnology, Inc.  
 ; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and  
 ; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer  
 ; FILE REFERENCE: 018501-012500US  
 ; CURRENT APPLICATION NUMBER: US/10/295,027  
 ; CURRENT FILING DATE: 2002-11-13  
 ; PRIOR APPLICATION NUMBER: US 09/663,733  
 ; PRIOR FILING DATE: 2000-09-15  
 ; PRIOR APPLICATION NUMBER: US 60/350,666  
 ; PRIOR FILING DATE: 2001-11-13  
 ; PRIOR APPLICATION NUMBER: US 60/335,394  
 ; PRIOR FILING DATE: 2001-11-15  
 ; PRIOR APPLICATION NUMBER: US 60/332,464  
 ; PRIOR FILING DATE: 2001-11-21  
 ; PRIOR APPLICATION NUMBER: US 60/334,393  
 ; PRIOR FILING DATE: 2001-11-29  
 ; PRIOR APPLICATION NUMBER: US 60/340,376  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: US 60/347,211  
 ; PRIOR FILING DATE: 2002-01-08  
 ; PRIOR APPLICATION NUMBER: US 60/347,349  
 ; PRIOR FILING DATE: 2002-01-10  
 ; PRIOR APPLICATION NUMBER: US 60/355,250  
 ; PRIOR FILING DATE: 2002-02-08  
 ; PRIOR APPLICATION NUMBER: US 60/356,714  
 ; PRIOR FILING DATE: 2002-02-13  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 1386  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 205  
 ; LENGTH: 461  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-295-027-205

Query Match 86.4%; Score 38; DB 15; Length 461;  
 Best Local Similarity 87.5%; Pred. No. 1e+03; 1; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KPSSPPE 8  
 Db 379 KPSSPPE 386

RESULT 41

US-10-188-832-97  
 ; Sequence 97, Application US/10188832  
 ; Publication No. US20040076955A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mack, David H.  
 ; APPLICANT: Aziz, Natasha  
 ; APPLICANT: Eos Biotechnology, Inc.  
 ; TITLE OF INVENTION: Methods of Diagnosis of Bladder Cancer, Compositions  
 ; TITLE OF INVENTION: and Methods of Screening for Modulators of Bladder  
 ; TITLE OF INVENTION: Cancer  
 ; FILE REFERENCE: 018501-002330US

US-10-188-832-97  
 ; CURRENT APPLICATION NUMBER: US/10/188,832  
 ; CURRENT FILING DATE: 2002-11-22  
 ; PRIOR APPLICATION NUMBER: US 60/302,814  
 ; PRIOR FILING DATE: 2001-07-03  
 ; PRIOR APPLICATION NUMBER: US 60/310,099  
 ; PRIOR FILING DATE: 2001-08-03  
 ; PRIOR APPLICATION NUMBER: US 60/343,705  
 ; PRIOR FILING DATE: 2001-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/350,666  
 ; PRIOR FILING DATE: 2001-11-13  
 ; PRIOR APPLICATION NUMBER: US 60/372,246  
 ; PRIOR FILING DATE: 2002-04-12  
 ; NUMBER OF SEQ ID NOS: 207  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 97  
 ; LENGTH: 461  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-188-832-97

Query Match 86.4%; Score 38; DB 16; Length 461;  
 Best Local Similarity 87.5%; Pred. No. 1e+03; 1; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KPSSPPE 8  
 Db 379 KPSSPPE 386

RESULT 42

US-09-972-035A-20  
 ; Sequence 20, Application US/09972035A  
 ; Patent No. US20020173622A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wettstein, Daniel A  
 ; APPLICANT: Morham, Scott  
 ; APPLICANT: Zavitz, Kenton  
 ; TITLE OF INVENTION: Tsg101-GAGp6 INTERACTION AND USE THEREOF  
 ; FILE REFERENCE: 1907.03  
 ; CURRENT APPLICATION NUMBER: US/09/972,035A  
 ; CURRENT FILING DATE: 2001-10-04  
 ; PRIOR APPLICATION NUMBER: US 60/276,259  
 ; PRIOR FILING DATE: 2001-03-14  
 ; NUMBER OF SEQ ID NOS: 32  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 20  
 ; LENGTH: 8  
 ; TYPE: PRT  
 ; ORGANISM: Human immunodeficiency virus  
 ; US-09-972-035A-20

Query Match 84.1%; Score 37; DB 9; Length 8;  
 Best Local Similarity 75.0%; Pred. No. 1e+06; 0; Indels 0; Gaps 0;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPE 8  
 Db 1 EPSAPPE 8

RESULT 43

US-10-223-172A-20  
 ; Sequence 20, Application US/10223172A  
 ; Publication No. US2003018444A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Myriad Genetics, Inc.  
 ; APPLICANT: Zavitz, Kenton  
 ; APPLICANT: Wettstein, Daniel Albert  
 ; APPLICANT: Morham, Scott  
 ; APPLICANT: Hobden, Adrian  
 ; TITLE OF INVENTION: COMPOSITION AND METHOD FOR TREATING HIV INFECTION  
 ; FILE REFERENCE: 5003.01  
 ; CURRENT APPLICATION NUMBER: US/10/223,172A

CURRENT FILING DATE: 2002-08-19  
PRIOR APPLICATION NUMBER: US 60/313,239  
PRIOR FILING DATE: 2001-08-18  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 20  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Human immunodeficiency virus  
US-10-223-172A-20

Query Match 84.1%; Score 37; DB 14; Length 8;  
Best Local Similarity 75.0%; Pred. No. 1e+06;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSPPEE 8  
Db 1 EPSAPPEE 8

RESULT 44  
US-10-224-999A-20  
Sequence 20, Application US/10224999A  
Publication No. US20030171318A1  
GENERAL INFORMATION:  
APPLICANT: Myriad Genetics, Inc.  
APPLICANT: Morham, Scott  
APPLICANT: Zavitz, Kenton  
APPLICANT: Hobden, Adrian  
TITLE OF INVENTION: Composition and Method for Treating Viral Infection  
FILE REFERENCE: 5004.01  
CURRENT APPLICATION NUMBER: US/10/224,999A  
CURRENT FILING DATE: 2003-03-03  
PRIOR APPLICATION NUMBER: US 60/313,695  
PRIOR FILING DATE: 2001-08-20  
NUMBER OF SEQ ID NOS: 3484  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 20  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Human immunodeficiency virus  
US-10-224-999A-20

Query Match 84.1%; Score 37; DB 14; Length 8;  
Best Local Similarity 75.0%; Pred. No. 1e+06;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSPPEE 8  
Db 1 EPSAPPEE 8

RESULT 45  
US-10-663-407-20  
Sequence 20, Application US/10663407  
Publication No. US20040109861A1  
GENERAL INFORMATION:  
APPLICANT: Myriad Genetics, Incorporated  
APPLICANT: Wetstein, Daniel A  
APPLICANT: Morham, Scott  
APPLICANT: Zavitz, Kenton  
TITLE OF INVENTION: TSG101-GAG INTERACTION AND USE THEREOF  
FILE REFERENCE: 1907.04-1  
CURRENT APPLICATION NUMBER: US/10/663,407  
CURRENT FILING DATE: 2003-09-15  
PRIOR APPLICATION NUMBER: PCT/US02/08146  
PRIOR FILING DATE: 2002-03-14  
PRIOR APPLICATION NUMBER: US 10/223,172  
PRIOR FILING DATE: 2002-08-19  
PRIOR APPLICATION NUMBER: US 10/224,999  
PRIOR FILING DATE: 2002-08-20  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatentIn version 3.2

SEQ ID NO 20  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Human immunodeficiency virus  
US-10-663-407-20

Query Match 84.1%; Score 37; DB 16; Length 8;  
Best Local Similarity 75.0%; Pred. No. 1e+06;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSPPEE 8  
Db 1 EPSAPPEE 8

RESULT 46  
US-09-972-035A-24  
Sequence 24, Application US/09972035A  
Patent No. US20020173622A1  
GENERAL INFORMATION:  
APPLICANT: Wetstein, Daniel A  
APPLICANT: Morham, Scott  
APPLICANT: Zavitz, Kenton  
TITLE OF INVENTION: TSG101-GAGp6 INTERACTION AND USE THEREOF  
FILE REFERENCE: 1907.03  
CURRENT APPLICATION NUMBER: US/09/972,035A  
CURRENT FILING DATE: 2001-10-04  
PRIOR APPLICATION NUMBER: US 60/276,259  
PRIOR FILING DATE: 2001-03-14  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 24  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Human immunodeficiency virus  
US-09-972-035A-24

Query Match 84.1%; Score 37; DB 9; Length 9;  
Best Local Similarity 75.0%; Pred. No. 1e+06;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSPPEE 8  
Db 2 EPSAPPEE 9

RESULT 47  
US-10-223-172A-24  
Sequence 24, Application US/10223172A  
Publication No. US20030138444A1  
GENERAL INFORMATION:  
APPLICANT: Myriad Genetics, Inc.  
APPLICANT: Zavitz, Kenton  
APPLICANT: Wetstein, Daniel Albert  
APPLICANT: Morham, Scott  
APPLICANT: Hobden, Adrian  
TITLE OF INVENTION: COMPOSITION AND METHOD FOR TREATING HIV INFECTION  
FILE REFERENCE: 5003.01  
CURRENT APPLICATION NUMBER: US/10/223,172A  
CURRENT FILING DATE: 2002-08-19  
PRIOR APPLICATION NUMBER: US 60/313,239  
PRIOR FILING DATE: 2001-08-18  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 24  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Human immunodeficiency virus  
US-10-223-172A-24

Query Match 84.1%; Score 37; DB 14; Length 9;  
Best Local Similarity 75.0%; Pred. No. 1e+06;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSPPEE 8  
:|:|:|  
Db 2 EPSAPPEE 9

## RESULT 48

US-10-224-999A-24  
; Sequence 24, Application US/10224999A  
; Publication No. US20030171318A1  
; GENERAL INFORMATION:  
; APPLICANT: Myriad Genetics, Inc.  
; APPLICANT: Morham, Scott  
; APPLICANT: Zavitz, Kenton  
; APPLICANT: Hobden, Adrian  
; TITLE OF INVENTION: Composition and Method for Treating Viral Infection  
; FILE REFERENCE: 5004.01  
; CURRENT APPLICATION NUMBER: US/10/224,999A  
; CURRENT FILING DATE: 2003-03-03  
; PRIOR APPLICATION NUMBER: US 60/313,695  
; PRIOR FILING DATE: 2001-08-20  
; NUMBER OF SEQ ID NOS: 3484  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 24  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus  
US-10-224-999A-24

Query Match 84.1%; Score 37; DB 14; Length 9;  
Best Local Similarity 75.0%; Pred. No. 1e+06; 0; Indels 0; Gaps 0;  
Matches 6; Conservative 2; Mismatches 0;

QY 1 KPSPPEE 8  
:|:|:|  
Db 2 EPSAPPEE 9

## RESULT 49

US-10-663-407-24  
; Sequence 24, Application US/10663407  
; Publication No. US20040109861A1  
; GENERAL INFORMATION:  
; APPLICANT: Myriad Genetics, Incorporated  
; APPLICANT: Wettstein, Daniel A  
; APPLICANT: Morham, Scott  
; APPLICANT: Zavitz, Kenton  
; TITLE OF INVENTION: TSG101-GAG INTERACTION AND USE THEREOF  
; FILE REFERENCE: 1907.04-1  
; CURRENT APPLICATION NUMBER: US/10/663,407  
; CURRENT FILING DATE: 2003-09-15  
; PRIOR APPLICATION NUMBER: PCT/US02/08146  
; PRIOR FILING DATE: 2002-03-14  
; PRIOR APPLICATION NUMBER: US 10/223,172  
; PRIOR FILING DATE: 2002-08-19  
; PRIOR APPLICATION NUMBER: US 10/224,999  
; PRIOR FILING DATE: 2002-08-20  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 24  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus  
US-10-663-407-24

Query Match 84.1%; Score 37; DB 16; Length 9;  
Best Local Similarity 75.0%; Pred. No. 1e+06; 0; Indels 0; Gaps 0;  
Matches 6; Conservative 2; Mismatches 0;

QY 1 KPSPPEE 8  
:|:|:|  
Db 2 EPSAPPEE 9

RESULT 50  
US-09-972-035A-26  
; Sequence 26, Application US/09972035A  
; Patent No. US20020173622A1  
; GENERAL INFORMATION:  
; APPLICANT: Wettstein, Daniel A  
; APPLICANT: Morham, Scott  
; APPLICANT: Zavitz, Kenton  
; TITLE OF INVENTION: Tsg101-GAGp6 INTERACTION AND USE THEREOF  
; FILE REFERENCE: 1907.03  
; CURRENT APPLICATION NUMBER: US/09/972,035A  
; CURRENT FILING DATE: 2001-10-04  
; PRIOR APPLICATION NUMBER: US 60/276,259  
; PRIOR FILING DATE: 2001-03-14  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 26  
; LENGTH: 10  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus  
US-09-972-035A-26

Query Match 84.1%; Score 37; DB 9; Length 10;  
Best Local Similarity 75.0%; Pred. No. 39; 0; Indels 0; Gaps 0;  
Matches 6; Conservative 2; Mismatches 0;

QY 1 KPSPPEE 8  
:|:|:|  
Db 3 EPSAPPEE 10

Search completed: June 18, 2004, 13:01:57  
Job time : 49 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2004, 12:52:03 ; Search time 55 Seconds  
(without alignments)  
41.098 Million cell updates/sec

Title: US-10-655-201-2

Perfect score: 44

Sequence: 1 KPSPPER 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 500 summaries

Database : A\_Geneseq.29yan04:\*

1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	100.0	8	2	AAW95057 Anti-Inv
2	44	100.0	8	5	ABH82274 Tumour an
3	44	100.0	8	5	ABH82273 Peptide C
4	44	100.0	8	7	ADC66464 Human utro
5	44	100.0	9	2	AAW95058 Urokinase
6	44	100.0	10	2	AAW95056 Urokinase
7	44	100.0	10	5	ABH82272 Plasminog
8	44	100.0	10	7	ADC66463 Human utro
9	44	100.0	21	1	AAH81672 Sequence
10	44	100.0	21	2	AAW22737 Linking s
11	44	100.0	96	5	AAH16550 Human uPA
12	44	100.0	143	5	AAH16549 Human uPA
13	44	100.0	208	2	AAW22747 Metastasi
14	44	100.0	259	2	AAH63140 Low molec
15	44	100.0	276	5	AAH16546 Human LHM
16	44	100.0	323	5	AAH16548 Human uPA
17	44	100.0	337	5	AAH16549 Human uPA
18	44	100.0	337	5	AAH16549 Human uPA
19	44	100.0	365	2	AAH68854 Delta 1-4
20	44	100.0	378	2	AAH16535 Human pro
21	44	100.0	386	2	AAH66266 Bifunctio
22	44	100.0	389	2	AAH16536 Human pro
23	44	100.0	390	2	AAH66245 Bifunctio
24	44	100.0	390	2	AAH66247 Bifunctio
25	44	100.0	392	2	AAH66260 Bifunctio

26	44	100.0	392	2	AAH66264 Bifunctio
27	44	100.0	392	2	AAH66255 Bifunctio
28	44	100.0	392	2	AAH66259 Bifunctio
29	44	100.0	392	2	AAH66261 Bifunctio
30	44	100.0	392	2	AAH66258 Bifunctio
31	44	100.0	392	2	AAH66263 Bifunctio
32	44	100.0	392	2	AAH66254 Bifunctio
33	44	100.0	392	2	AAH66256 Bifunctio
34	44	100.0	392	2	AAH66257 Bifunctio
35	44	100.0	393	2	AAH66251 Bifunctio
36	44	100.0	393	2	AAH66253 Bifunctio
37	44	100.0	393	2	AAH66249 Bifunctio
38	44	100.0	393	2	AAH66252 Bifunctio
39	44	100.0	393	2	AAH66244 Bifunctio
40	44	100.0	393	2	AAH66250 Bifunctio
41	44	100.0	393	2	AAH99597 ChimERIC
42	44	100.0	393	2	AAH99596 ChimERIC
43	44	100.0	393	2	AAH99885 M36: fibr
44	44	100.0	395	2	AAH47902 Pro-uroki
45	44	100.0	395	2	AAH66265 Bifunctio
46	44	100.0	395	2	AAH66262 Bifunctio
47	44	100.0	396	2	AAH66246 Bifunctio
48	44	100.0	397	2	AAH11828 FB-FB-UK
49	44	100.0	397	2	AAH66248 Bifunctio
50	44	100.0	401	2	AAH13637 Human pro
51	44	100.0	410	2	AAH23794 Prourokin
52	44	100.0	411	1	AAH50871 Sequence
53	44	100.0	411	1	AAH91664 Sequence
54	44	100.0	411	1	AAH96146 Sequence
55	44	100.0	411	1	AAH93589 Amino aci
56	44	100.0	411	1	AAH93504 Sequence
57	44	100.0	411	2	AAH66244 Urokinase
58	44	100.0	411	2	AAH07904 Human pro
59	44	100.0	411	2	AAH07902 Human pro
60	44	100.0	411	2	AAH07903 Human pro
61	44	100.0	411	2	AAH05117 UK-S3 as
62	44	100.0	411	2	AAH13634 Human nat
63	44	100.0	411	2	AAH10057 Pro-uroki
64	44	100.0	411	2	AAH10058 Pro-uroki
65	44	100.0	411	2	AAH10056 Pro-uroki
66	44	100.0	411	2	AAH22098 Pro-UK, 3
67	44	100.0	411	2	AAH34584 Mutant hu
68	44	100.0	411	2	AAH47956 PUK G16A
69	44	100.0	411	2	AAH47966 PUK G16A
70	44	100.0	411	2	AAH47962 PUK G38A
71	44	100.0	411	2	AAH47963 PUK G53A
72	44	100.0	411	2	AAH40225 PUK, 2/19
73	44	100.0	411	2	AAH47965 PUK N32P
74	44	100.0	411	2	AAH47958 PUK S26T
75	44	100.0	411	2	AAH47961 PUK G38A
76	44	100.0	411	2	AAH47959 PUK G32P
77	44	100.0	411	2	AAH47960 PUK P34A
78	44	100.0	411	2	AAH47964 PUK N32P
79	44	100.0	411	2	AAH47957 PUK Y24A
80	44	100.0	411	2	AAH47956 PUK G16A
81	44	100.0	411	2	AAH63008 Pro-uroki
82	44	100.0	411	2	AAH62989 Pro-uroki
83	44	100.0	411	2	AAH62994 Pro-uroki
84	44	100.0	411	2	AAH63003 Pro-uroki
85	44	100.0	411	2	AAH63001 Pro-uroki
86	44	100.0	411	2	AAH63006 Pro-uroki
87	44	100.0	411	2	AAH62992 Pro-uroki
88	44	100.0	411	2	AAH63002 Pro-uroki
89	44	100.0	411	2	AAH63007 Pro-uroki
90	44	100.0	411	2	AAH62995 Pro-uroki
91	44	100.0	411	2	AAH63000 Pro-uroki
92	44	100.0	411	2	AAH63004 Pro-uroki
93	44	100.0	411	2	AAH62996 Pro-uroki
94	44	100.0	411	2	AAH63005 Pro-uroki
95	44	100.0	411	2	AAH62991 Pro-uroki
96	44	100.0	411	2	AAH62998 Pro-uroki
97	44	100.0	411	2	AAH62993 Pro-uroki
98	44	100.0	411	2	AAH92926 Pro-uroki

99	44	100.0	411	2	AAW19709	Aw19709	Pro-uroki	172	44	100.0	434	2	AAW20537	Aw20537	Amidated
100	44	100.0	411	2	AAW39343	Aw39343	Human pro	173	44	100.0	434	2	AAW20538	Aw20538	Amidated
101	44	100.0	411	2	AAW42284	Aw42284	Human pro	174	44	100.0	436	2	AAW20536	Aw20536	Amidated
102	44	100.0	411	3	AAW92836	Aw92836	Urokinase	175	44	100.0	473	1	AAW60797	Aw60797	Interfero
103	44	100.0	411	3	AAW78343	Aw78343	Human pro	176	44	93.2	8	2	AAW95052	Aw95052	Urokinase
104	44	100.0	411	4	AAW20489	Aw20489	Human pro	177	41	93.2	8	5	AAW82277	Aw82277	Urokinase
105	44	100.0	411	4	AAW74797	Aw74797	Prourokin	178	41	90.9	8	7	ADCC6470	AdC6470	Human uro
106	44	100.0	411	5	AAW16544	Aw16544	Human uro	179	40	90.9	380	2	AAW05433	Aw05433	CPA-P2 Hy
107	44	100.0	411	5	AAW13269	Aw13269	Human pro	180	40	90.9	410	2	AAW23795	Aw23795	Prourokin
108	44	100.0	411	7	ADCC72159	AdC72159	Human uro	181	39	88.6	7	2	AAW95063	Aw95063	Urokinase
109	44	100.0	411	7	ADCC65977	AdC65977	Human sin	182	39	88.6	7	2	AAW95054	Aw95054	Urokinase
110	44	100.0	412	2	AAW10334	Aw10334	Recombinda	183	39	88.6	7	5	AAW82279	Aw82279	Tumour an
111	44	100.0	421	2	AAW10173	Aw10173	Human pro	184	39	88.6	7	5	AAW82278	Aw82278	Tumour an
112	44	100.0	421	2	AAW10172	Aw10172	Human pro	185	39	88.6	7	7	ADCC6467	AdC6467	Human uro
113	44	100.0	424	2	AAW24579	Aw24579	Inhibitor	186	39	88.6	7	7	ADCC6466	AdC6466	Human uro
114	44	100.0	425	2	AAW24578	Aw24578	Human pro	187	39	88.6	8	7	ADCC6472	AdC6472	Human uro
115	44	100.0	430	2	AAW24578	Aw24578	Inhibitor	188	39	88.6	11	2	AAW95059	Aw95059	Urokinase
116	44	100.0	430	3	AAW78341	Aw78341	Human pre	189	39	88.6	11	2	AAW82280	Aw82280	Tumour an
117	44	100.0	431	1	AAW30041	Aw30041	Sequence	190	39	88.6	27	2	AAW05105	Aw05105	Plasminog
118	44	100.0	431	1	AAW50503	Aw50503	Sequence	191	39	88.6	380	2	AAW22504	Aw22504	[GARSQ] -
119	44	100.0	431	1	AAW50114	Aw50114	Sequence	192	39	88.6	411	2	AAW80428	Aw80428	Sequence
120	44	100.0	431	1	AAW60783	Aw60783	Human uro	193	39	88.6	411	2	AAW04185	Aw04185	Plasminog
121	44	100.0	431	1	AAW60674	Aw60674	Modified	194	39	88.6	476	2	AAW22503	Aw22503	[GARSQ] -
122	44	100.0	431	1	AAW70258	Aw70258	Sequence	195	39	88.6	919	4	AAW07234	Aw07234	Novel hum
123	44	100.0	431	1	AAW71491	Aw71491	Modified	196	39	88.6	1178	7	ADCC6468	AdC6468	Letuce E
124	44	100.0	431	1	AAW71663	Aw71663	Modified	197	38	86.4	8	2	AAW95060	Aw95060	Urokinase
125	44	100.0	431	1	AAW71698	Aw71698	Modified	198	38	86.4	8	2	AAW95061	Aw95061	Urokinase
126	44	100.0	431	1	AAW71699	Aw71699	Modified	199	38	86.4	8	5	AAW82276	Aw82276	Tumour an
127	44	100.0	431	1	AAW70250	Aw70250	Sequence	200	38	86.4	8	5	AAW82275	Aw82275	Tumour an
128	44	100.0	431	1	AAW80430	Aw80430	Deduced A	201	38	86.4	8	7	ADCC6469	AdC6469	Human uro
129	44	100.0	431	1	AAW81204	Aw81204	Pro-uroki	202	38	86.4	166	4	ADCC6468	AdC6468	Human uro
130	44	100.0	431	1	AAW91886	Aw91886	Sequence	203	38	86.4	166	4	AAW18264	Aw18264	Peptide #
131	44	100.0	431	1	AAW92119	Aw92119	Natural h	204	38	86.4	166	4	AAW15177	Aw15177	Peptide #
132	44	100.0	431	1	AAW92119	Aw92119	Non-glyco	205	38	86.4	166	4	AAW15170	Aw15170	Peptide #
133	44	100.0	431	2	AAW07112	Aw07112	Human pro	206	38	86.4	166	4	AAW37296	Aw37296	Peptide #
134	44	100.0	431	2	AAW04253	Aw04253	Human pro	207	38	86.4	166	4	AAW27634	Aw27634	Peptide #
135	44	100.0	431	2	AAW63141	Aw63141	Full leng	208	38	86.4	166	4	AAW30756	Aw30756	Peptide #
136	44	100.0	431	2	AAW47903	Aw47903	Pro-uroki	209	38	86.4	166	4	AAW829001	Aw829001	Peptide #
137	44	100.0	431	2	AAW33199	Aw33199	Human uro	210	38	86.4	166	4	AAW32047	Aw32047	Peptide #
138	44	100.0	431	2	AAW33198	Aw33198	Human uro	211	38	86.4	166	4	AAW82585	Aw82585	Protein #
139	44	100.0	431	2	AAW33121	Aw33121	Human uro	212	38	86.4	166	4	AAW67343	Aw67343	Human bon
140	44	100.0	431	3	AAW50865	Aw50865	Human uro	213	38	86.4	166	4	AAW70434	Aw70434	Human bon
141	44	100.0	431	3	AAW95951	Aw95951	Human pla	214	38	86.4	166	4	AAW57998	Aw57998	Human bva
142	44	100.0	431	3	AAW84605	Aw84605	Amho aci	215	38	86.4	166	4	AAW49006	Aw49006	Human liv
143	44	100.0	431	5	AAW17128	Aw17128	Human uro	216	38	86.4	166	4	AAW52114	Aw52114	Human liv
144	44	100.0	431	5	AAW92332	Aw92332	Human uro	217	38	86.4	166	4	AAW05878	Aw05878	Peptide #
145	44	100.0	431	5	AAW92328	Aw92328	Human pla	218	38	86.4	166	4	AAW02915	Aw02915	Peptide #
146	44	100.0	431	5	AAW92327	Aw92327	Human pla	219	38	86.4	166	5	ABG40068	AbG40068	Human pep
147	44	100.0	431	5	AAW92342	Aw92342	Human pla	220	38	86.4	166	5	ABG36981	AbG36981	Human pep
148	44	100.0	431	5	AAW92240	Aw92240	Human pla	221	38	86.4	166	5	ABW898525	AbW898525	Zinc iron
149	44	100.0	431	5	AAW92338	Aw92338	Human pla	222	38	86.4	461	5	AAW99908	Aw99908	Human 647
150	44	100.0	431	5	AAW92331	Aw92331	Human pla	223	38	86.4	461	6	ADW4577	AdW4577	Human pro
151	44	100.0	431	5	AAW92329	Aw92329	Human pla	224	38	86.4	461	6	ABW48190	AbW48190	Human bla
152	44	100.0	431	5	AAW92330	Aw92330	Human pla	225	38	86.4	461	6	ABW56642	AbW56642	Lung canc
153	44	100.0	431	5	AAW92334	Aw92334	Human pla	226	38	86.4	477	6	ABW56642	AbW56642	Lung canc
154	44	100.0	431	5	AAW92339	Aw92339	Human pla	227	37	84.1	8	5	ABW54774	AbW54774	Peptide 1
155	44	100.0	431	5	AAW92243	Aw92243	Human pla	228	37	84.1	8	6	ABW65621	AbW65621	HIV GAG P
156	44	100.0	431	5	AAW92233	Aw92233	Human pla	229	37	84.1	8	6	ABW65621	AbW65621	HIV GAG P
157	44	100.0	431	5	AAW92235	Aw92235	Human pla	230	37	84.1	9	5	ABW65625	AbW65625	HIV GAG P
158	44	100.0	431	5	AAW92241	Aw92241	Human pla	231	37	84.1	9	6	ABW65625	AbW65625	HIV GAG P
159	44	100.0	431	5	AAW92236	Aw92236	Human pla	232	37	84.1	9	6	ABW65625	AbW65625	HIV GAG P
160	44	100.0	431	6	AAW92241	Aw92241	Human uro	233	37	84.1	10	6	ABW65630	AbW65630	HIV GAG P
161	44	100.0	431	6	AAW92241	Aw92241	Human uro	234	37	84.1	10	6	ABW65630	AbW65630	HIV GAG P
162	44	100.0	431	6	AAW92241	Aw92241	Human uro	235	37	84.1	10	6	ABW65630	AbW65630	HIV GAG P
163	44	100.0	431	6	AAW92241	Aw92241	Human uro	236	37	84.1	10	6	ABW65630	AbW65630	HIV GAG P
164	44	100.0	431	6	AAW92241	Aw92241	Human uro	237	37	84.1	10	6	ABW65630	AbW65630	HIV GAG P
165	44	100.0	431	6	AAW92241	Aw92241	Human uro	238	37	84.1	10	6	ABW65630	AbW65630	HIV GAG P
166	44	100.0	431	6	AAW92241	Aw92241	Human uro	239	37	84.1	10	6	ABW65630	AbW65630	HIV GAG P
167	44	100.0	431	7	AAW92241	Aw92241	Human uro	240	37	84.1	10	6	ABW65630	AbW65630	HIV GAG P
168	44	100.0	431	7	AAW92241	Aw92241	Human uro	241	37	84.1	10	6	ABW65630	AbW65630	HIV GAG P
169	44	100.0	431	7	AAW92241	Aw92241	Human uro	242	37	84.1	10	6	ABW65630	AbW65630	HIV GAG P
170	44	100.0	431	7	AAW92241	Aw92241	Human uro	243	37	84.1	10	6	ABW65630	AbW65630	HIV GAG P
171	44	100.0	431	7	AAW92241	Aw92241	Human uro	244	37	84.1	10	6	ABW65630	AbW65630	HIV GAG P

245	36	81.8	10	6	ABP6534	Abp96534	HTV GAG P	318	36	81.8	1089	6	ABR59317	Abt59317	Human sec
246	36	81.8	56	2	AA74180	Aay74180	Human pro	319	36	81.8	1089	6	ABO09379	Aboc09379	Human sec
247	36	81.8	52	2	AA74400	Aay74400	Human pro	320	36	81.8	1089	6	ABO19243	Aboc19243	Novel hum
248	36	81.8	84	6	ABR58380	Abt58380	Human NOV	321	36	81.8	1089	6	ABO11261	Aboc11261	Human sec
249	36	81.8	111	3	AAO3717	Aag03717	Human sec	322	36	81.8	1089	6	ABR66879	Abt66879	Human sec
250	36	81.8	134	4	AAO10357	Aao10357	Human pol	323	36	81.8	1089	6	ABO16092	Aboc16092	Human sec
251	36	81.8	263	2	ABR60326	Abb60326	Drosophila	324	36	81.8	1089	6	ABO13798	Aboc13798	Human sec
252	36	81.8	362	4	AAW97364	Aaw97364	Amino aci	325	36	81.8	1089	6	ABU65701	Abu65701	Human sec
253	36	81.8	362	4	AAU0305	Aae0305	Human pig	326	36	81.8	1089	6	ABO07549	Aboc07549	Human PRO
254	36	81.8	362	4	AAU08652	Aau08652	Human ant	327	36	81.8	1089	6	ABO03736	Aboc03736	Human sec
255	36	81.8	362	5	ABR08433	Abb08433	SLED poly	328	36	81.8	1089	6	ABR67184	Abt67184	Human sec
256	36	81.8	362	5	ABR04587	Abbo4587	Human ant	329	36	81.8	1089	6	ABO15787	Aboc15787	Human sec
257	36	81.8	411	5	AA316551	Aae16551	Human uro	330	36	81.8	1089	6	ABU56068	Abu56068	Human sec
258	36	81.8	418	2	AA44860	Aar44860	Sequence	331	36	81.8	1089	6	ABU95341	Abu95341	Novel hum
259	36	81.8	418	2	AA490287	Aar90287	Pigment e	332	36	81.8	1089	6	ABU65396	Abu65396	Human PRO
260	36	81.8	418	4	AAE10306	Aae10306	Human pig	333	36	81.8	1089	6	ABU71244	Abu71244	Human PRO
261	36	81.8	418	5	ABR81091	Abb81091	Human pig	334	36	81.8	1089	6	ABO070854	Aboc070854	Human PRO
262	36	81.8	418	6	ABR672122	Abg72122	Human pig	335	36	81.8	1089	6	ABR70095	Abt70095	Human sec
263	36	81.8	418	6	ABR58379	Abt58379	Human NOV	336	36	81.8	1089	6	ABR69428	Abt69428	Human sec
264	36	81.8	418	6	AAE38136	Aae38136	Human ret	337	36	81.8	1089	6	ABO01569	Aboc01569	Human PRO
265	36	81.8	418	7	ABU10031	Abu10031	Human pig	338	36	81.8	1089	6	ABU81371	Abu81371	Human PRO
266	36	81.8	418	7	ABR59089	Abb59089	Human ret	339	36	81.8	1089	6	ABR60168	Abt60168	Human sec
267	36	81.8	494	4	AA393227	Abb70299	Drosophila	340	36	81.8	1089	6	ABR67903	Abt67903	Human sec
268	36	81.8	574	3	AA393227	Aab32227	Human pro	341	36	81.8	1089	6	ABR65291	Abt65291	Human sec
269	36	81.8	671	3	AA324257	Aab24257	Arabidops	342	36	81.8	1089	6	ABR68513	Abt68513	Human sec
270	36	81.8	671	4	AAU00463	Aau00463	Arabidops	343	36	81.8	1089	6	ABR71925	Abt71925	Human PRO
271	36	81.8	672	6	AAE37052	Aae37052	Human gen	344	36	81.8	1089	6	ABU85405	Abu85405	Human PRO
272	36	81.8	672	6	AAE37066	Aae37066	Human gen	345	36	81.8	1089	6	ABU89095	Abu89095	Human sec
273	36	81.8	672	6	AAE37276	Aae37276	Human gen	346	36	81.8	1089	6	ABU83175	Abu83175	Human sec
274	36	81.8	747	4	AA84251	Aae4251	Human cyc	347	36	81.8	1089	6	ABU95031	Abu95031	Novel hum
275	36	81.8	851	7	ADCG4335	Adcg4335	KIRA1196.	348	36	81.8	1089	6	ABU90579	Abu90579	Novel hum
276	36	81.8	880	7	ADD14150	Add14150	Human src	349	36	81.8	1089	6	ABU84090	Abu84090	Human sec
277	36	81.8	896	4	AAW4252	Aaw4252	Human EST	350	36	81.8	1089	6	ABU93741	Abu93741	Novel hum
278	36	81.8	988	4	AA84252	Abd84252	Human aei	351	36	81.8	1089	6	ABR64986	Abt64986	Human sec
279	36	81.8	988	3	AA939369	Aay9369	Human PRO	352	36	81.8	1089	6	ABR68818	Abt68818	Human sec
280	36	81.8	1089	4	AAU29156	Aau29156	Protein O	353	36	81.8	1089	6	ABO06634	Aboc06634	Human sec
281	36	81.8	1089	4	AAU29156	Aau29156	Human PRO	354	36	81.8	1089	6	ABR99179	Abt99179	Human sec
282	36	81.8	1089	6	ABU85832	Abu85832	Human PRO	355	36	81.8	1089	6	ABU57063	Abu57063	Human PRO
283	36	81.8	1089	6	ABU88080	Abu88080	Novel hum	356	36	81.8	1089	6	ABU86015	Abu86015	Novel hum
284	36	81.8	1089	6	ABU84395	Abu84395	Human sec	357	36	81.8	1089	6	ABU82302	Abu82302	Novel hum
285	36	81.8	1089	6	ABR66269	Abt66269	Human sec	358	36	81.8	1089	6	ABU87313	Abu87313	Human PRO
286	36	81.8	1089	6	ABR65659	Abt65659	Human sec	359	36	81.8	1089	6	ABU83785	Abu83785	Human sec
287	36	81.8	1089	6	ABU99599	Abu99599	Human sec	360	36	81.8	1089	6	ABO08159	Aboc08159	Human PRO
288	36	81.8	1089	6	ABU82838	Abu82838	Human PRO	361	36	81.8	1089	6	ABU81870	Abu81870	Novel hum
289	36	81.8	1089	6	ABU89959	Abu89959	Novel hum	362	36	81.8	1089	6	ABU66034	Abu66034	Novel hum
290	36	81.8	1089	6	ABR68208	Abt68208	Human sec	363	36	81.8	1089	6	ABR59863	Abt59863	Human sec
291	36	81.8	1089	6	ABU96261	Abu96261	Novel hum	364	36	81.8	1089	6	ABU94051	Abu94051	Novel hum
292	36	81.8	1089	6	ABU92692	Abu92692	Human sec	365	36	81.8	1089	6	ABU99904	Abu99904	Novel hum
293	36	81.8	1089	6	ABO08769	Aboc08769	Human sec	366	36	81.8	1089	6	ABR66574	Abt66574	Human sec
294	36	81.8	1089	6	ABO02821	Aboc02821	Human sec	367	36	81.8	1089	6	ABR90992	Abt90992	Human sec
295	36	81.8	1089	6	ABR74975	Abt74975	Human sec	368	36	81.8	1089	6	ABU94419	Abu94419	Human PRO
296	36	81.8	1089	6	ABR94737	Abt94737	Human sec	369	36	81.8	1089	6	ABU79301	Abu79301	Human PRO
297	36	81.8	1089	6	ABU85710	Abu85710	Human PRO	370	36	81.8	1089	6	ABU86630	Abu86630	Human sec
298	36	81.8	1089	6	ABU98870	Abu98870	Novel hum	371	36	81.8	1089	6	ABU86935	Abu86935	Novel hum
299	36	81.8	1089	6	ABU98085	Abu98085	Novel hum	372	36	81.8	1089	6	ABU94724	Abu94724	Human PRO
300	36	81.8	1089	6	ABU91791	Abu91791	Novel hum	373	36	81.8	1089	6	ABO04651	Aboc04651	Human PRO
301	36	81.8	1089	6	ABU89484	Abu89484	Human PRO	374	36	81.8	1089	6	ABR70400	Abt70400	Human sec
302	36	81.8	1089	6	ABU86325	Abu86325	Human sec	375	36	81.8	1089	6	ABU98565	Abu98565	Human PRO
303	36	81.8	1089	6	ABU67538	Abu67538	Human sec	376	36	81.8	1089	6	ABR65965	Abt65965	Human sec
304	36	81.8	1089	6	ABU80566	Abu80566	Human PRO	377	36	81.8	1089	6	ABR64681	Abt64681	Human sec
305	36	81.8	1089	6	ABR9484	Abt9484	Human sec	378	36	81.8	1089	6	ABU79606	Abu79606	Human PRO
306	36	81.8	1089	6	ABR98874	Abt98874	Human sec	379	36	81.8	1089	6	ABU92997	Abu92997	Human sec
307	36	81.8	1089	6	ABO16397	Aboc16397	Human sec	380	36	81.8	1089	6	ABU95956	Abu95956	Human PRO
308	36	81.8	1089	6	ABR92257	Abt92257	Human sec	381	36	81.8	1089	6	ABU91176	Abu91176	Novel hum
309	36	81.8	1089	6	ABO18938	Aboc18938	Human sec	382	36	81.8	1089	6	ABO90269	Aboc90269	Novel hum
310	36	81.8	1089	6	ABR78359	Abt78359	Human sec	383	36	81.8	1089	6	ABO09684	Aboc09684	Human sec
311	36	81.8	1089	6	ABU85095	Abu85095	Human sec	384	36	81.8	1089	6	ABO10956	Aboc10956	Human sec
312	36	81.8	1089	6	ABO00234	Aboc00234	Novel hum	385	36	81.8	1089	6	ABR71010	Abt71010	Human sec
313	36	81.8	1089	6	ABO11566	Aboc11566	Human sec	386	36	81.8	1089	6	ABU87618	Abu87618	Human PRO
314	36	81.8	1089	6	ABO02211	Aboc02211	Human sec	387	36	81.8	1089	6	ABU91486	Abu91486	Human PRO
315	36	81.8	1089	6	ABU88785	Abu88785	Novel hum	388	36	81.8	1089	6	ABU84700	Abu84700	Human sec
316	36	81.8	1089	6	ABU83480	Abu83480	Human sec	389	36	81.8	1089	6	ABR69790	Abt69790	Human sec
317	36	81.8	1089	6	ABO06281	Aboc06281	Novel hum	390	36	81.8	1089	6	ABU80167	Abu80167	Human PRO

391	36	81.8	1089	6	ABU93436	Abu93436 Human PRO
392	36	81.8	1089	6	ABO09989	AbO09989 Human sec
393	36	81.8	1089	6	ABO09074	AbO09074 Human sec
394	36	81.8	1089	6	ABU10642	AbU10642 Human sec
395	36	81.8	1089	6	ABU95651	AbU95651 Human PRO
396	36	81.8	1089	6	ABU96860	AbU96860 Novel hum
397	36	81.8	1089	6	ABR70705	AbR70705 Human sec
398	36	81.8	1089	6	ABO05056	AbO05056 Novel hum
399	36	81.8	1089	6	ABO08464	AbO08464 Human sec
400	36	81.8	1089	6	ABO05671	AbO05671 Human sec
401	36	81.8	1089	6	ABR74060	AbR74060 Human sec
402	36	81.8	1089	6	ABR95652	AbR95652 Human sec
403	36	81.8	1089	6	ABR80949	AbR80949 Human sec
404	36	81.8	1089	6	ABR81254	AbR81254 Human sec
405	36	81.8	1089	6	ABM00950	AbM00950 Human sec
406	36	81.8	1089	6	ABR88552	AbR88552 Human sec
407	36	81.8	1089	6	ABM77373	AbM77373 Human sec
408	36	81.8	1089	6	ABO28857	AbO28857 Human sec
409	36	81.8	1089	6	ABO31602	AbO31602 Human sec
410	36	81.8	1089	6	ABM08019	AbM08019 Human sec
411	36	81.8	1089	6	ABO40499	AbO40499 Human sec
412	36	81.8	1089	6	ABO35924	AbO35924 Human PRO
413	36	81.8	1089	6	ABO44063	AbO44063 Human PRO
414	36	81.8	1089	6	ADA78018	AdA78018 Human sec
415	36	81.8	1089	6	ABM24858	AbM24858 Human sec
416	36	81.8	1089	6	ABO03126	AbO03126 Human sec
417	36	81.8	1089	6	ABR90382	AbR90382 Human sec
418	36	81.8	1089	6	ABM17296	AbM17296 Human sec
419	36	81.8	1089	6	ABR95042	AbR95042 Human sec
420	36	81.8	1089	6	ABR95347	AbR95347 Human sec
421	36	81.8	1089	6	ABO21585	AbO21585 Human sec
422	36	81.8	1089	6	ABR97849	AbR97849 Human sec
423	36	81.8	1089	6	ABR87637	AbR87637 Human sec
424	36	81.8	1089	6	ABM77678	AbM77678 Human sec
425	36	81.8	1089	6	ABM27908	AbM27908 Human sec
426	36	81.8	1089	6	ABM06189	AbM06189 Human sec
427	36	81.8	1089	6	ABM03695	AbM03695 Human sec
428	36	81.8	1089	6	ABM35146	AbM35146 Human sec
429	36	81.8	1089	6	ABM26383	AbM26383 Human sec
430	36	81.8	1089	6	ABO48165	AbO48165 Human sec
431	36	81.8	1089	6	ABR92907	AbR92907 Human sec
432	36	81.8	1089	6	ABO24668	AbO24668 Human sec
433	36	81.8	1089	6	ABM11679	AbM11679 Human sec
434	36	81.8	1089	6	ABM02780	AbM02780 Human sec
435	36	81.8	1089	6	ABM16076	AbM16076 Human sec
436	36	81.8	1089	6	ABO27637	AbO27637 Human sec
437	36	81.8	1089	6	ABM29128	AbM29128 Human sec
438	36	81.8	1089	6	ABM07104	AbM07104 Human sec
439	36	81.8	1089	6	ABM21198	AbM21198 Human sec
440	36	81.8	1089	6	ABM09544	AbM09544 Human sec
441	36	81.8	1089	6	ABO41414	AbO41414 Human sec
442	36	81.8	1089	6	ABO36229	AbO36229 Human PRO
443	36	81.8	1089	6	ABO43758	AbO43758 Human PRO
444	36	81.8	1089	6	ABM76458	AbM76458 Human sec
445	36	81.8	1089	6	ABM76154	AbM76154 Human sec
446	36	81.8	1089	6	ABM25773	AbM25773 Human sec
447	36	81.8	1089	6	ABM26078	AbM26078 Human sec
448	36	81.8	1089	6	ABO03431	AbO03431 Human sec
449	36	81.8	1089	6	ABO02516	AbO02516 Human sec
450	36	81.8	1089	6	ABR90687	AbR90687 Human sec
451	36	81.8	1089	6	ABR73785	AbR73785 Human sec
452	36	81.8	1089	6	ABO17007	AbO17007 Human sec
453	36	81.8	1089	6	ABR94432	AbR94432 Human sec
454	36	81.8	1089	6	ABR75939	AbR75939 Human sec
455	36	81.8	1089	6	ABR71315	AbR71315 Human sec
456	36	81.8	1089	6	ABR93212	AbR93212 Human sec
457	36	81.8	1089	6	ABR93517	AbR93517 Human sec
458	36	81.8	1089	6	ABR87942	AbR87942 Human sec
459	36	81.8	1089	6	ABO33611	AbO33611 Novel hum
460	36	81.8	1089	6	ABO27942	AbO27942 Human sec
461	36	81.8	1089	6	ABO30077	AbO30077 Human sec
462	36	81.8	1089	6	ABO33286	AbO33286 Human PRO
463	36	81.8	1089	6	ABM04974	AbM04974 Human sec

464	36	81.8	1089	6	ABM08934	AbM08934 Human sec
465	36	81.8	1089	6	ABO36534	AbO36534 Human sec
466	36	81.8	1089	6	ABO35619	AbO35619 Human PRO
467	36	81.8	1089	6	ABO39584	AbO39584 Human sec
468	36	81.8	1089	6	ABM10459	AbM10459 Human sec
469	36	81.8	1089	6	ABM11984	AbM11984 Human sec
470	36	81.8	1089	6	ABO52130	AbO52130 Human PRO
471	36	81.8	1089	6	ABO52435	AbO52435 Human PRO
472	36	81.8	1089	6	ABO23753	AbO23753 Human sec
473	36	81.8	1089	6	ABR97239	AbR97239 Human sec
474	36	81.8	1089	6	ABR87027	AbR87027 Human sec
475	36	81.8	1089	6	ABM11069	AbM11069 Human sec
476	36	81.8	1089	6	ABM28213	AbM28213 Human sec
477	36	81.8	1089	6	ABO32212	AbO32212 Human sec
478	36	81.8	1089	6	ABM15339	AbM15339 Human sec
479	36	81.8	1089	6	ABM06494	AbM06494 Human sec
480	36	81.8	1089	6	ABM04305	AbM04305 Human sec
481	36	81.8	1089	6	ABM22418	AbM22418 Human sec
482	36	81.8	1089	6	ABM07714	AbM07714 Human sec
483	36	81.8	1089	6	ABO40804	AbO40804 Human sec
484	36	81.8	1089	6	ABM35451	AbM35451 Human sec
485	36	81.8	1089	6	ABM33214	AbM33214 Human sec
486	36	81.8	1089	6	ABO52740	AbO52740 Human PRO
487	36	81.8	1089	6	ABO50300	AbO50300 Human sec
488	36	81.8	1089	6	ABU99294	AbU99294 Human sec
489	36	81.8	1089	6	ABO04346	AbO04346 Human sec
490	36	81.8	1089	6	ABO05976	AbO05976 Human sec
491	36	81.8	1089	6	ABM18516	AbM18516 Human sec
492	36	81.8	1089	6	ABR97544	AbR97544 Human sec
493	36	81.8	1089	6	ABR80644	AbR80644 Human sec
494	36	81.8	1089	6	ABR801255	AbR801255 Human sec
495	36	81.8	1089	6	ABR88857	AbR88857 Human sec
496	36	81.8	1089	6	ABM13509	AbM13509 Human sec
497	36	81.8	1089	6	ABM20893	AbM20893 Human sec
498	36	81.8	1089	6	ABO42024	AbO42024 Human sec
499	36	81.8	1089	6	ABO42634	AbO42634 Human sec
500	36	81.8	1089	7	ADE50392	AdE50392 Human sec

ALIGNMENTS

RESULT 1

AAW95057

ID AAW95057 standard; peptide; 8 AA.

XX

AC AAW95057;

DT 19-MAY-1999 (first entry)

XX

DE Anti-invasive and anti-angiogenic urokinase peptide fragment.

XX

KW Urokinase; anti-invasive; anti-angiogenic; variant; chemical derivative; affinity ligand; pharmaceutical; inhibition; tumour invasion; metastasis; peptidomimetic; cell migration; proliferation; angiogenesis; fibrosis; atherosclerosis; post-Balloon angioplasty vascular restenosis; neointima; vascular trauma; vascular graft restenosis; inflammatory; lung; scarring; wound healing; psoriasis; venous thrombosis.

XX

OS Synthetic.

XX

FT Key

FT Modified-site 1 Location/Qualifiers

FT /note= "N-terminal acetylation"

FT Modified-site 8 /note= "C-terminal amide"

FT

PN WO9905263-AI.

XX

PD 04-FEB-1999.

XX

PF 24-JUL-1998; 98WO-US015437.

XX







```

DT      19-MAY-1999 (first entry)
XX      Urokinase peptide fragment A14.
DE
XX      Urokinase; anti-invasive; anti-angiogenic; variant; chemical derivative;
XX      affinity ligand; pharmaceutical; inhibition; tumour invasion; metastasis;
XX      peptidomimetic; cell migration; proliferation; angiogenesis; fibrosis;
XX      atherosclerosis; post-balloon angioplasty vascular restenosis; neointima;
XX      vascular trauma; vascular graft restenosis; inflammatory; lung; scarring;
XX      wound healing; psoriasis; venous thrombosis.
XX      Synthetic.
XX      OS
XX      WO9905263-A1.
XX      PD
XX      04-FEB-1999.
XX      PF
XX      24-JUL-1998; 98WO-US015437.
XX      PR
XX      25-JUL-1997; 97US-00900327.
XX      PA
XX      (ANGS-) ANGSTROM PHARM INC.
XX      PI
XX      Jones TR, Mazar AP;
XX      WPI: 1999-142921/12.
XX      DR
XX      New urokinase peptides and derivatives with anti-invasive and anti-
XX      PT angiogenic activity - useful for treating diseases or conditions
XX      PT associated with undesired cell migration, invasion, migration-induced
XX      PT proliferation, or angiogenesis.
XX      PS
XX      Example 2; Page 44; 73pp; English.
XX      CC
XX      The invention relates to an anti-invasive and anti-angiogenic peptide
XX      CC compound. The compound has a sequence corresponding to that shown in
XX      CC AAM95057 or a substitution variant, addition variant or other chemical
XX      CC derivative of that peptide. The peptide or its variant or a derivative
XX      CC can be capped or uncapped. The peptides are useful for in vivo or in
XX      CC vitro prognostic and diagnostic applications, e.g. as labelled peptides
XX      CC to detect a binding site for the peptide on a surface or in the interior
XX      CC of a cell (see AAM95057 for detailed uses of the peptide compound, its
XX      CC variants/derivatives). The present sequence represents a urokinase
XX      CC peptide fragment tested along with the peptide of the invention for anti-
XX      CC invasive activity
XX      SO
XX      Sequence 9 AA:
XX
XX      Query Match 100.0%; Score 44; DB 2; Length 9;
XX      Best Local Similarity 100.0%; Pred. No. 1.4e+06;
XX      Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
XX
Oy      1 KPSSPPEE 8
      1 |||||
      1 KPSSPPEE 8
Db
      1 KPSSPPEE 8

RESULT 6
AAM95056
ID      AAM95056 standard; peptide; 10 AA.
XX
XX      AAM95056;
XX      DT
XX      19-MAY-1999 (first entry)
XX      DE
XX      Urokinase peptide fragment.
XX      Urokinase; anti-invasive; anti-angiogenic; variant; chemical derivative;
XX      affinity ligand; pharmaceutical; inhibition; tumour invasion; metastasis;
XX      peptidomimetic; cell migration; proliferation; angiogenesis; fibrosis;
XX      atherosclerosis; post-balloon angioplasty vascular restenosis; neointima;
XX      vascular trauma; vascular graft restenosis; inflammatory; lung; scarring;
XX      wound healing; psoriasis; venous thrombosis.

```

```

XX OS Synthetic.
XX PN W09905263-A1.
XX PD 04-FEB-1999.
XX PF 24-JUL-1998; 98WO-US015437.
XX PR 25-JUL-1997; 97US-00900327.
XX PA (ANGS-) ANGSTROM PHARM INC.
XX PI Jones TR, Mazar AP;
XX DR WPI; 1999-142921/12.
XX PT New urokinase peptides and derivatives with anti-invasive and anti-
XX PT angiogenic activity - useful for treating diseases or conditions
XX PT associated with undesired cell migration, invasion, migration-induced
XX PT proliferation, or angiogenesis.
XX PS Example 2; Page 43; 73pp; English.
XX CC The invention relates to an anti-invasive and anti-angiogenic peptide
XX CC compound. The compound has a sequence corresponding to that shown in
XX CC AA95057 or a substitution variant, addition variant or other chemical
XX CC derivative of that peptide. The peptide or its variant or a derivative
XX CC can be capped or uncapped. The peptides are useful for in vivo or in
XX CC vitro prognostic and diagnostic applications, e.g. as labelled peptides
XX CC to detect a binding site for the peptide on a surface or in the interior
XX CC of a cell (see AA95057 for detailed uses of the peptide compound, its
XX CC variants/derivatives). The present sequence represents a urokinase
XX CC peptide fragment tested along with the peptide of the invention for anti-
XX CC invasive activity
XX SQ Sequence 10 AA;

Query Match 100.0%; Score 44; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8
   |||||
DB 1 KPSSPPEE 8

RESULT 7
ABB82272
ID ABB82272 standard; peptide; 10 AA.
AC ABB82272;
XX 08-JAN-2003 (first entry)
DE Plasminogen activator uPA tryptic phosphopeptide (residues 136-145).
XX Plasminogen; uPA; tumour; cytosolic; ophthalmological; antidiabetic;
XX antiproliferative; vulnery; antiinflammatory; vasotropic; gynaecological;
XX antitumour; cytotoxic; chromolytic.
XX Unidentified.
XX OS
XX WO20026985-A2.
XX 12-SEP-2002.
XX 22-JAN-2002; 2002WO-US001560.
XX 19-JAN-2001; 2001US-0262422P.
XX (ANGS-) ANGSTROM PHARM INC.
XX

```



FT Region 1..7  
 FT /note= "t-PA residues 955-975"  
 FT 8..21  
 FT /note= "u-PA-II residues 134-147"  
 XX  
 XX WO8080451-A.  
 XX  
 PD 03-NOV-1988.  
 XX  
 XX 28-APR-1988; 88WO-NL000020.  
 XX  
 XX 29-APR-1987; 87NL-00001021.  
 XX  
 XX (BLOE-) STICHT LAB BLOEDTFA.  
 XX (NERO-) NEDERLAND RODE KRUI.  
 XX  
 PI Pannekoek H;  
 XX  
 DR WPI; 1988-322771/45.  
 DR N-PSDB; AAN81682.  
 XX  
 PT Mutant plasminogen activator protein(s) - contg. a t-PA portion including  
 the H chain and a u-PA portion including the B chain.  
 XX  
 PS Disclosure; Fig 1; 39pp; English.  
 XX  
 CC t-PA/u-PA substitution-mutant proteins t-PA::u-PA-I and t-PA::U-PA-II are  
 CC claimed. The mutant is much more effective in plasminogen activation in  
 CC the presence of fibrin and is more resistant to inactivation by a PA  
 CC inhibitor. It is used in pharmaceutical compsn. having an effect on blood  
 CC clotting and/or on fibrinolysis. (Updated on 25-MAR-2003 to correct PA  
 CC field.)  
 XX  
 SQ Sequence 21 AA;

Query Match 100.0%; Score 44; DB 1; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 6.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8  
 DB 10 KPSSPPEE 17

RESULT 10  
 AAW22737  
 ID AAW22737 standard; peptide; 21 AA.  
 XX  
 AC AAW22737;  
 XX  
 XX

DT 12-MAR-1998 (first entry)  
 XX  
 XX

DE Linking sequence from metastasis inhibitor.  
 XX  
 XX

KW Chimeric; human; urinary trypsin; inhibitor; HI-8; cancer; metastasis;  
 G-domain; urokinase; prevention; leukaemia; lymphoma.  
 XX  
 XX

OS Synthetic.  
 XX  
 XX

PN WO9725422-A1.  
 XX  
 XX

PD 17-JUL-1997.  
 XX  
 XX

PF 06-JAN-1997; 97WO-JP000008.  
 XX  
 XX

PR 08-JAN-1996; 96JP-00001059.  
 XX  
 XX

PA (NISP) NISSIN FOOD PROD CO LTD.  
 XX  
 XX

PI Kobayashi H, Terao T, Sugino D, Okushima M;  
 XX  
 XX

WPI; 1997-372862/34.  
 XX  
 XX

PT Chimeric protein which inhibits development of metastases in cancer -  
 PT contains urinary trypsin inhibitor carboxy-terminal domain linked to  
 PT urokinase G-domain.  
 XX  
 XX

PS Claim 2; Page 68; 97pp; Japanese.  
 XX  
 XX

CC A novel chimeric protein contains the carboxy-terminal domain of human  
 CC urinary trypsin inhibitor (HI-8), which inhibits cancer cell metastasis,  
 CC linked to a peptide containing the G-domain of urokinase (AAW22742),  
 CC which specifically binds the excess urokinase receptor expressed in  
 CC cancer cells. The chimeric protein has the amino-terminal AAW22734, the  
 CC carboxy-terminal AAW22735 and a linking sequence selected from AAW22736-  
 CC 39 or partial sequences derived from these, specifically AAW3130-63. The  
 CC chimeric protein may also have additional amino-terminal sequences  
 CC selected from AAW22740 or 9 partial sequences derived from this, and/or  
 CC additional carboxy-terminal sequences selected from AAW22743 or 10  
 CC partial sequences derived from this. The chimeric protein can be used to  
 CC prevent metastasis in, e.g. cancer of the lung, kidney, pancreas,  
 CC stomach, colon, rectum, ovary, uterus, brain, skin, muscle, breast or  
 CC prostate, and in leukaemia or lymphoma  
 XX  
 XX

SQ Sequence 21 AA;

Query Match 100.0%; Score 44; DB 2; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 6.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8  
 DB 5 KPSSPPEE 12

RESULT 11  
 AAE16550  
 ID AAE16550 standard; protein; 96 AA.  
 XX  
 AC AAE16550;  
 XX  
 XX

DT 09-APR-2002 (first entry)  
 XX  
 XX

DE Human uPA kringle and connecting peptide.  
 XX  
 XX

KW Human; urokinase-type plasminogen activator; uPA; therapy; hypertension;  
 KW stroke; hypotension; atherosclerosis; heart attack; thrombotic disorder;  
 KW microvascular occlusion; angiogenic disorder; pulmonary fibrosis; asthma;  
 KW tumour cell metastasis; glaucoma; diabetic retinopathy; wound healing;  
 KW clotting disorder; uterine contraction disorder; respiratory disease;  
 KW adult respiratory distress syndrome; male impotence.  
 XX  
 XX

OS Homo sapiens.  
 XX  
 XX

PN WO200197752-A2.  
 XX  
 XX

PD 27-DEC-2001.  
 XX  
 XX

PF 13-JUN-2001; 2001WO-US018976.  
 XX  
 XX

PR 20-JUN-2000; 2000US-0212874P.  
 XX  
 XX

PA (UYPR-) UNIV PENNSYLVANIA.  
 XX  
 XX

PI Cines DB, Higazi AA;  
 XX  
 XX

DR WPI; 2002-122240/16.  
 XX  
 XX

DR N-PSDB; AAD27083.  
 XX  
 XX

PT Composition for modulating muscle cell and tissue contractility for  
 PT treating atherosclerosis, asthma, hypertension, glaucoma, impotence,  
 PT comprising domains from urokinase-type plasminogen activator.  
 XX  
 XX

Claim 25; Fig 11; 117pp; English.

The invention relates to a composition comprising one or more domains of

CC urokinase-type plasminogen activator (uPA). The composition is used to  
 CC modulate the contractility and angiogenic activity of a mammalian muscle,  
 CC endothelial cell or tissue. The composition is used for treating stroke,  
 CC hypertension, atherosclerosis, heart attack, microvascular  
 CC occlusions, thrombotic microangiopathies, surgically induced thrombotic  
 CC disorders, angiogenic disorders, pulmonary fibrosis, asthma, tumour cell  
 CC invasion, tumour cell angiogenesis, tumour cell metastasis, glaucoma,  
 CC diabetic retinopathy, wound healing, clotting disorder, uterine  
 CC contraction disorder, male impotence, respiratory disease or condition  
 CC such as asthma, adult respiratory distress syndrome, primary pulmonary  
 CC hypertension, microvascular thrombotic occlusion, and a disorder  
 CC associated with chronic intrapulmonary fibrin formation. The present  
 CC sequence is human urokinase-type plasminogen activator (uPA) kringite and  
 CC connecting peptide

QY Sequence 96 AA;  
 SQ

Query Match 100.0%; Score 44; DB 5; Length 96;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPSSPPEE 8  
 DB 89 KPSSPPEE 96

RESULT 12  
 AAE16549  
 ID AAE16549 standard; protein; 143 AA.  
 AC AAE16549;  
 XX  
 DT 09-APR-2002 (first entry)  
 XX

DE Human uPA amino terminal fragment (ATF) and connecting peptide.  
 XX  
 XX Human; urokinase-type plasminogen activator; uPA; therapy; hypertension;  
 KW stroke; hypotension; atherosclerosis; heart attack; thrombotic disorder;  
 KW microvascular occlusion; angiogenic disorder; pulmonary fibrosis; asthma;  
 KW tumour cell metastasis; glaucoma; diabetic retinopathy; wound healing;  
 KW clotting disorder; uterine contraction disorder; respiratory disease;  
 KW adult respiratory distress syndrome; amino terminal fragment; ATF;  
 KW male impotence.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX MO200197752-A2.  
 PN  
 XX  
 XX 27-DEC-2001.  
 PD  
 XX  
 XX 13-JUN-2001; 2001WO-US018976.  
 PF  
 XX  
 XX 20-JUN-2000; 2000US-0212874P.  
 PR  
 XX  
 XX (UYFE-) UNIV PENNSYLVANIA.  
 PA  
 XX  
 XX Cines DB, Higazi AA;  
 PI  
 XX  
 XX WPI; 2002-122240/16.  
 DR  
 XX  
 XX N-PSDB; AAD27082.  
 DR

PT Composition for modulating muscle cell and tissue contractility for  
 PT treating atherosclerosis, asthma, hypertension, glaucoma, impotence,  
 PT comprising domains from urokinase-type plasminogen activator.  
 XX  
 XX Claim 24; Fig 1H; 117pp; English.  
 PS  
 XX  
 XX The invention relates to a composition comprising one or more domains of  
 CC urokinase-type plasminogen activator (uPA). The composition is used to  
 CC modulate the contractility and angiogenic activity of a mammalian muscle,  
 CC endothelial cell or tissue. The composition is used for treating stroke,  
 CC hypertension, atherosclerosis, heart attack, microvascular  
 CC occlusions, thrombotic microangiopathies, surgically induced thrombotic  
 CC disorders, angiogenic disorders, pulmonary fibrosis, asthma, tumour cell  
 CC invasion, tumour cell angiogenesis, tumour cell metastasis, glaucoma,  
 CC diabetic retinopathy, wound healing, clotting disorder, uterine  
 CC contraction disorder, male impotence, respiratory disease or condition  
 CC such as asthma, adult respiratory distress syndrome, primary pulmonary  
 CC hypertension, microvascular thrombotic occlusion, and a disorder  
 CC associated with chronic intrapulmonary fibrin formation. The present  
 CC sequence is human urokinase-type plasminogen activator (uPA) amino  
 CC terminal fragment (ATF) and connecting peptide

CC disorders, angiogenic disorders, pulmonary fibrosis, asthma, tumour cell  
 CC invasion, tumour cell angiogenesis, tumour cell metastasis, glaucoma,  
 CC diabetic retinopathy, wound healing, clotting disorder, uterine  
 CC contraction disorder, male impotence, respiratory disease or condition  
 CC such as asthma, adult respiratory distress syndrome, primary pulmonary  
 CC hypertension, microvascular thrombotic occlusion, and a disorder  
 CC associated with chronic intrapulmonary fibrin formation. The present  
 CC sequence is human urokinase-type plasminogen activator (uPA) amino  
 CC terminal fragment (ATF) and connecting peptide

QY Sequence 143 AA;  
 SQ

Query Match 100.0%; Score 44; DB 5; Length 143;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPSSPPEE 8  
 DB 136 KPSSPPEE 143

RESULT 13  
 AAM22747  
 ID AAM22747 standard; protein; 208 AA.  
 XX  
 XX AAM22747;  
 AC  
 XX  
 XX 12-MAR-1998 (first entry)  
 DT  
 XX  
 XX Metastasis inhibitor.  
 DE  
 XX  
 XX Chimeric; human; urinary trypsin; inhibitor; HI-8; cancer; metastasis;  
 KW G-domain; urokinase; prevention; leukaemia; lymphoma.  
 KW  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX MO9725422-A1.  
 PN  
 XX  
 XX 17-JUL-1997.  
 PD  
 XX  
 XX 06-JAN-1997; 97MO-JP000008.  
 PF  
 XX  
 XX 08-JAN-1996; 96JP-00001059.  
 PR  
 XX  
 XX (NISP) NISSIN FOOD PROD CO LTD.  
 PA  
 XX  
 XX Kobayashi H, Terao T, Sugino D, Okushima M;  
 PI  
 XX  
 XX WPI; 1997-372862/34.  
 DR  
 XX  
 XX N-PSDB; AAT75155.  
 DR

PT Chimeric protein which inhibits development of metastases in cancer -  
 PT containing urinary trypsin inhibitor carboxy-terminal domain linked to  
 PT urokinase G-domain.  
 XX  
 XX Claim 19; Page 59-60; 97pp; Japanese.  
 PS  
 XX  
 XX The present sequence is a novel chimeric protein, which contains the  
 CC carboxy-terminal domain of human urinary trypsin inhibitor (HI-8), which  
 CC inhibits cancer cell metastasis, linked to a peptide containing the G-  
 CC domain of urokinase, which specifically binds the excess urokinase  
 CC receptor expressed in cancer cells. The chimeric protein can be used to  
 CC prevent metastasis in, e.g. cancer of the lung, kidney, pancreas,  
 CC stomach, colon, rectum, ovary, uterus, brain, skin, muscle, breast or  
 CC prostate, and in leukemia or lymphoma  
 CC  
 XX  
 XX Sequence 208 AA;  
 SQ

Query Match 100.0%; Score 44; DB 2; Length 208;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPSSPPEE 8

Db |||||  
137 KPSSPPEE 144

RESULT 14  
AAE63140  
ID AAE63140 standard; protein; 259 AA.

XX  
AC AAE63140;

XX  
DT 25-MAR-2003 (revised)  
DT 02-JUN-1995 (first entry)

XX  
DE Low molecular weight urokinase bioactive protein.

XX  
KM Urokinase bioactive glycoprotein; cardiovascular diseases;  
KW pulmonary embolism.

XX  
OS Homo sapiens.

XX  
PN EP620279-A1.

XX  
PD 19-OCT-1994.

XX  
PF 14-APR-1983; 94EP-00104777.

XX  
PR 15-APR-1982; 82US-00368773.

XX  
PR 14-MAR-1983; 83US-00474930.

XX  
PR 14-APR-1983; 83EP-00103629.

XX  
PA (GETH ) GENENTECH INC.

XX  
PI Heyneker HL, Holmes WE, Vehar GA;

XX  
DR WPI; 1994-318362/40.

XX  
DR N-PSDB; AAO73481.

XX  
PT Prodn. of human urokinase glycoprotein - using a recombinant expression  
PT system used for the treatment of vascular diseases or conditions.

XX  
PS Claim 6; Fig 2; 41pp; English.

XX  
CC AAQ73481 is the cDNA sequence which encodes AAE63140 a low molecular  
CC weight 33000 dalton urokinase (UK) bioactive protein. This cDNA was used  
CC in the construction of a plasmid capable of transforming either yeast or  
CC vertebrate cells, enabling them to produce the 33000 dalton UK bioactive  
CC protein. The UK glycoprotein produced could then be used in the treatment  
CC of cardiovascular diseases, including pulmonary embolism. The UK produced  
CC using this method had the advantage of a specific activity towards fibrin  
CC and extant thrombi, not demonstrated previously with UK isolated from  
CC natural sources. (Updated on 25-MAR-2003 to correct PN field.) (Updated  
CC on 25-MAR-2003 to correct PF field.) (Updated on 25-MAR-2003 to correct  
CC PR field.)

XX  
SQ Sequence 259 AA;

Query Match 100.0%; Score 44; DB 2; Length 259;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8  
Db |||||  
4 KPSSPPEE 11

RESULT 15  
AAE16546  
ID AAE16546 standard; protein; 276 AA.

XX  
AC AAE16546;

XX  
DT 09-APR-2002 (first entry)

XX

DE Human LMW-urokinase-type plasminogen activator (uPA) protein.

XX  
KM Human; urokinase-type plasminogen activator; uPA; therapy; hypertension;  
KM stroke; hypotension; atherosclerosis; heart attack; thrombotic disorder;  
KM microvascular occlusion; angiogenic disorder; pulmonary fibrosis; asthma;  
KM tumour cell metastasis; glaucoma; diabetic retinopathy; wound healing;  
KM clotting disorder; uterine contraction disorder; respiratory disease;  
KM male impotence; adult respiratory distress syndrome; LMW-uPA.

XX  
OS Homo sapiens.

XX  
PN WO200197752-A2.

XX  
PD 27-DEC-2001.

XX  
PF 13-JUN-2001; 2001MO-US018976.

XX  
PR 20-JUN-2000; 2000US-0212874P.

XX  
PA (UYPE-) UNIV PENNSYLVANIA.

XX  
PI Cines DB, Higazi AA;

XX  
DR WPI; 2002-122240/16.

XX  
DR N-PSDB; AAD27079.

XX  
PT Composition for modulating muscle cell and tissue contractility for  
PT treating atherosclerosis, asthma, hypertension, glaucoma, impotence,  
PT comprising domains from urokinase-type plasminogen activator.

XX  
PS Claim 5; Fig 1E; 117pp; English.

XX  
CC The invention relates to a composition comprising one or more domains of  
CC urokinase-type plasminogen activator (uPA). The composition is used to  
CC modulate the contractility and angiogenic activity of a mammalian muscle,  
CC endothelial cell or tissue. The composition is used for treating stroke,  
CC hypotension, hypertension, atherosclerosis, heart attack, microvascular  
CC occlusions, thrombotic microangiopathies, surgically induced thrombotic  
CC disorders, angiogenic disorders, pulmonary fibrosis, asthma, tumour cell  
CC invasion, tumour cell angiogenesis, tumour cell metastasis, glaucoma,  
CC diabetic retinopathy, wound healing, clotting disorder, uterine  
CC contraction disorder, male impotence, respiratory disease or condition  
CC such as asthma, adult respiratory distress syndrome, primary pulmonary  
CC hypertension, microvascular thrombotic occlusion, and a disorder  
CC associated with chronic intrapulmonary fibrin formation. The present  
CC sequence is human LMW-urokinase-type plasminogen activator (uPA) protein

XX  
SQ Sequence 276 AA;

Query Match 100.0%; Score 44; DB 5; Length 276;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8  
Db |||||  
1 KPSSPPEE 8

RESULT 16  
AAE16548  
ID AAE16548 standard; protein; 323 AA.

XX  
AC AAE16548;

XX  
DT 09-APR-2002 (first entry)

XX  
DE Human uPA deltalakringle-secuPA and deltalakringle-tcuPA mutant.

XX  
KM Human; urokinase-type plasminogen activator; uPA; therapy; hypertension;  
KM stroke; hypotension; atherosclerosis; heart attack; thrombotic disorder;  
KM microvascular occlusion; angiogenic disorder; pulmonary fibrosis; asthma;  
KM tumour cell metastasis; glaucoma; diabetic retinopathy; wound healing;  
KM clotting disorder; uterine contraction disorder; respiratory disease;





PD 03-JAN-2002.  
 XX 07-JUN-2001; 2001WO-US018569.  
 PF 07-JUN-2000; 2000US-0209467P.  
 PR (HUMA-) HUMAN GENOME SCI INC.  
 PA Blase CE, Rosen CA;  
 PI WPI; 2002-147878/19.  
 XX N-PSDB; ABQ54872.  
 DR Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
 PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian  
 PT cancer), immune disorders, cardiovascular disorders and neurological  
 PT diseases.  
 XX Claim 11; SEQ ID NO 2927; 2922pp; English.  
 XX The invention relates to 2175 novel human ovarian antigens (ABP41054-  
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
 CC to the sequences of the invention. The invention additionally relates to  
 CC recombinant vectors and host cells comprising human ovarian antigen  
 CC polynucleotides, antibodies against human ovarian antigens, and the use  
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
 CC treating, prognosing or preventing various ovarian and/or breast-related  
 CC disorders. Such conditions include ovarian cancer and breast cancer, and  
 CC metastatic tumors of ovarian or breast origin, reproductive system  
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine  
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
 CC vaginitis), immune disorders (e.g., congenital and acquired  
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
 CC respiratory disorders, neurological disorders, gastrointestinal disorders  
 CC and urinary system disorders. Ovarian antigen polypeptides and  
 CC polynucleotides may also be used in screening for compounds which  
 CC modulate ovarian antigen expression or activity. The polynucleotides may  
 CC further be used for gene therapy, chromosome mapping, in the  
 CC identification of individuals and in forensic analysis, and the  
 CC polypeptides may be used as food additives or to prepare antibodies  
 CC useful in disease diagnosis, drug targeting and phenotyping. The present  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 337 AA:  
 SO  
 Query Match 100.0%; Score 44; DB 5; Length 337;  
 Best Local Similarity 100.0%; Pred. No. 85;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KPSSPPEE 8  
 DB 162 KPSSPPEE 169  
 RESULT 19  
 AAR68854  
 ID AAR68854 standard; protein; 365 AA.  
 XX  
 AC AAR68854;  
 XX  
 DT 16-OCT-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 22-NOV-1995 (first entry)  
 XX  
 DE Delta 1-46 urokinase.

KW Human; des-epidermal growth factor homologous plasminogen activator; uPA;  
 KW liver membrane; reduced affinity; EGF homologous; thrombolysis;  
 KW thrombolytic; increased half-life; urokinase.  
 XX  
 OS Homo sapiens; (engineered).  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 1  
 FT /note= "amino acids 1-46 of wild-type urokinase have been  
 FT deleted"  
 XX  
 PN US5376547-A.  
 PD 27-DEC-1994.  
 XX  
 PF 29-JUN-1988; 88US-00150267.  
 XX  
 PR 30-JAN-1987; 87US-00008795.  
 XX  
 PA (AMHP) AMERICAN HOME PROD CORP.  
 XX  
 PI Hung PP, Lee SL, Kalyan NK;  
 PT WPI; 1995-043464/06.  
 XX  
 PT New modified plasminogen activator cpds. - having regions removed to  
 PT reduce affinity for liver membranes and increase circulation half-life.  
 XX  
 PS Claim 1; Page 7; 26pp; English.  
 XX  
 CC Amino acid residues 1-46 contain the EGF region of human urokinase.  
 CC Deletion of this region results in a plasminogen activator with reduced  
 CC affinity for liver cell membranes; the mutant protein is not cleared from  
 CC the circulation as rapidly as is wild-type tPA. The specification only  
 CC gives the sequence around the deletion and not the full-length sequence  
 CC of "delta 1-46 urokinase"; the sequence in AAR68854 has been obtained by  
 CC amending a previously disclosed wild-type human urokinase sequence (from  
 CC W09501427) according to the description given in Example 3. (Updated on  
 CC 25-MAR-2003 to correct PF field.) (Updated on 16-OCT-2003 to standardise  
 CC OS field)  
 CC  
 XX Sequence 365 AA;  
 SO  
 Query Match 100.0%; Score 44; DB 2; Length 365;  
 Best Local Similarity 100.0%; Pred. No. 92;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KPSSPPEE 8  
 DB 90 KPSSPPEE 97  
 RESULT 20  
 AAM13635  
 ID AAM13635 standard; protein; 378 AA.  
 XX  
 AC AAM13635;  
 XX  
 DT 04-JUN-1997 (first entry)  
 DT  
 DE Human prourokinase variant lacking entire EGF domain.  
 XX  
 KW Human; prourokinase; hPUK; variant; half-life; increase; EGF;  
 KW epidermal growth factor domain; deletion; thrombolysis; fibrinolysis.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT 1..9  
 FT /note= "residues 1-9 of native hPUK"  
 FT 10..378  
 FT /note= "residues 43-411 of native hPUK"

FT Misc-difference 144  
 FT /note= "corresponds to TAC codon"  
 XX  
 XX EP398361-A.  
 XX  
 PD 22-NOV-1990.  
 XX  
 PF 18-MAY-1990; 90EP-00109472.  
 XX  
 PR 18-MAY-1989; 89JP-00126433.  
 PR 22-FEB-1990; 90JP-00042020.  
 XX  
 PA (GREC ) GREEN CROSS CORP.  
 XX  
 PI Tanabe T, Amatsuji Y, Kasai S, Hirose M, Morita M, Kawabe H;  
 PI Airmura H;  
 XX  
 DR WPI, 1990-350146/47.  
 DR N-PSDB; AAT61672.  
 XX  
 PT Human pro-urokinase variants - deficient in loop regions of epidermal  
 PT growth factor, showing long blood half-life, as fibrinolytic agent.  
 XX  
 PS Claim 1; Page; 22pp; English.  
 XX  
 CC New variants of human prourokinase (hPUK) comprise a hPUK deficient in  
 CC (1) at least part of the first loop region of the epidermal growth factor  
 CC (EGF) domain; (ii) at least part of the first loop and at least part of  
 CC the second loop; or (iii) at least part of the third loop. The hPUK  
 CC variants show an increased blood half-life comparable to that of the  
 CC whole EGF domain-deficient hPUK variant and urokinase while retaining the  
 CC same properties as those of hPUK. They have potent thrombolytic activity  
 CC and very little tendency to cause spontaneous bleeding. The present  
 CC sequence represents a specific variant of hPUK which lacks the entire EGF  
 CC domain; the sequence does not appear in the specification and has been  
 CC created using the wild-type hPUK sequence and the junction sequence after  
 CC deletion, both of which are given (in Fig 1 and in Fig 2(3),  
 CC respectively)  
 CC  
 SQ Sequence 378 AA;  
 Query Match 100.0%; Score 44; DB 2; Length 378;  
 Best Local Similarity 100.0%; Pred. No. 95;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KPSSPPEE 8  
 Db 103 KPSSPPEE 110  
 RESULT 21  
 AAR6266  
 ID AAR6266 standard; protein; 386 AA.  
 AC AAR6266;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 17-AUG-1995 (first entry)  
 XX  
 DE Bifunctional urokinase variant M33.  
 XX  
 KW fibrinolysis; thrombin inhibition; thrombolytic; anti-thrombotic;  
 KW urokinase; variant; mutein.  
 XX  
 OS Synthetic.  
 XX  
 FH Key  
 FT Region  
 FT Disulfide-bond 4. .85  
 FT Disulfide-bond 25. .67  
 FT Disulfide-bond 56. .80

FT Disulfide-bond 102. .233  
 FT Disulfide-bond 143. .159  
 FT Disulfide-bond 151. .222  
 FT Disulfide-bond 247. .316  
 FT Disulfide-bond 279. .295  
 FT Disulfide-bond 306. .334  
 FT Disulfide-bond 365. .366  
 FT Region  
 FT /label= X1  
 FT /note= "peptide bond"  
 FT 366. .386  
 FT /label= Y1  
 PN DE4323754-CI.  
 XX  
 PD 01-DEC-1994.  
 XX  
 PF 15-JUL-1993; 93DE-04323754.  
 XX  
 PR 15-JUL-1993; 93DE-04323754.  
 XX  
 PA (CHEF ) GRUENENTHAL GMBH.  
 XX  
 PI Steffens GJ, Whendt S, Schneider J, Heinzel-Wieland R;  
 PI Saunders DJ;  
 XX  
 DR WPI, 1995-015191/03.  
 XX  
 PT New bifunctional urokinase derivs and related plasmids - with improved  
 PT fibrinolytic and thrombin inhibiting activities, for treating cardiac and  
 PT cerebral infarct, pulmonary embolism, etc.  
 XX  
 PS Example 1; Page 11 and Fig 1; 34pp; German.  
 XX  
 CC Bifunctional urokinase derivatives corresponding to the formula M4-X1-Y1  
 CC are claimed (see features table). Sequences AAR6244-R66266 are specific  
 CC examples of such derivs, which have both improved fibrinolytic and  
 CC thrombin-inhibiting activities, compared to known plasminogen activators  
 CC or thrombin inhibitors. The proteins are useful as thrombolytic agents,  
 CC e.g. for treatment of arterial occlusions, deep vein thrombosis, cardiac  
 CC and cerebral infarction and pulmonary embolism. (Updated on 25-MAR-2003  
 CC to correct FN field.)  
 CC  
 SQ Sequence 386 AA;  
 Query Match 100.0%; Score 44; DB 2; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 97;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KPSSPPEE 8  
 Db 90 KPSSPPEE 97  
 RESULT 22  
 AAM13636  
 ID AAM13636 standard; protein; 389 AA.  
 AC AAM13636;  
 XX  
 DT 04-JUN-1997 (first entry)  
 XX  
 DE Human prourokinase variant lacking EGF domain loops 1 and 2.  
 XX  
 KW Human prourokinase; hPUK; variant; half-life; increase; EGF;  
 KW epidermal growth factor domain; deletion; thrombolytic; fibrinolysis.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 FH Key  
 FT Region  
 FT Disulfide-bond 1. .10  
 FT Disulfide-bond 11. .389

FT /note= "residues 33-411 of native hPUK"  
 FT Misc-difference 155  
 FT /note= "corresponds to TAC codon"  
 XX  
 XX EP398361-A.  
 XX  
 XX 22-NOV-1990.  
 XX  
 XX 18-MAY-1990; 90EP-00109472.  
 XX  
 XX 18-MAY-1989; 89JP-00126433.  
 XX  
 XX 22-FEB-1990; 90JP-00042020.  
 XX  
 XX (GREG ) GREEN CROSS CORP.  
 XX  
 XX Tanabe T, Amatsuji Y, Kasai S, Hirose M, Morita M, Kawabe H,  
 XX  
 XX Aikura H;  
 XX  
 XX WPI; 1990-350146/47.  
 XX  
 XX N-PSDB; AAT61673.  
 XX  
 XX Human pro-urokinase variants - deficient in loop regions of epidermal  
 XX growth factor, showing long blood half-life, as fibrinolytic agent.  
 XX  
 XX Claim 6; Page; 22pp; English.  
 XX  
 XX New variants of human prourokinase (hPUK) comprise a hPUK deficient in  
 XX (i) at least part of the first loop region of the epidermal growth factor  
 XX (EGF) domain; (ii) at least part of the first loop and at least part of  
 XX the second loop; or (iii) at least part of the third loop. The hPUK  
 XX variants show an increased blood half-life comparable to that of the  
 XX whole EGF domain-deficient hPUK variant and urokinase while retaining the  
 XX same properties as those of hPUK. They have potent thrombolytic activity  
 XX and very little tendency to cause spontaneous bleeding. The present  
 XX sequence represents a specific variant of hPUK which lacks loops 1 and 2  
 XX of the EGF domain; the sequence does not appear in the specification and  
 XX has been created using the wild-type hPUK sequence and the junction  
 XX sequence after deletion, both of which are given (in Fig 1 and on page 8,  
 XX respectively)  
 XX  
 XX SQ Sequence 389 AA;  
 XX  
 XX Query Match 100.0%; Score 44; DB 2; Length 389;  
 XX Best Local Similarity 100.0%; Pred. No. 98;  
 XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 XX QY 1 KPSSPPE 8  
 XX |||||  
 XX 114 KPSSPPE 121  
 XX  
 XX RESULT 23  
 XX AAR66245  
 XX ID AAR66245 standard; protein; 390 AA.  
 XX  
 XX AAR66245;  
 XX  
 XX 25-MAR-2003 (revised)  
 XX DT 17-AUG-1995 (first entry)  
 XX  
 XX Bifunctional urokinase variant M12.  
 XX  
 XX fibrinolysis; thrombin inhibition; thrombolytic; anti-thrombotic;  
 XX urokinase; variant; mutein.  
 XX  
 XX Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 XX FH 1..365  
 XX Region /label= M4  
 XX FT /note= "unglycosylated prourokinase(Ser47-Leu411)"  
 XX FT Disulfide-bond 4..85  
 XX FT Disulfide-bond 25..67

FT Disulfide-bond 56..80  
 FT Disulfide-bond 102..233  
 FT Disulfide-bond 143..159  
 FT Disulfide-bond 151..222  
 FT Disulfide-bond 247..316  
 FT Disulfide-bond 279..295  
 FT Disulfide-bond 306..334  
 FT Disulfide-bond 366..371  
 FT Region /label= X1  
 FT 372..390  
 FT Region /label= Y1  
 XX  
 XX DE4323754-C1.  
 XX  
 XX 01-DEC-1994.  
 XX  
 XX 15-JUL-1993; 93DE-04323754.  
 XX  
 XX 15-JUL-1993; 93DE-04323754.  
 XX  
 XX (CHER ) GRUNENTHAL GMBH.  
 XX  
 XX Steffens GJ, Whendt S, Schneider J, Heinzel-Wieland R;  
 XX Saunders DJ;  
 XX  
 XX WPI; 1995-015191/03.  
 XX  
 XX New bifunctional urokinase derivs and related plasmids - with improved  
 XX fibrinolytic and thrombin inhibiting activities, for treating cardiac and  
 XX cerebral infarct, pulmonary embolism, etc.  
 XX  
 XX Example 1; Page 10 and Fig 1; 34pp; German.  
 XX  
 XX Bifunctional urokinase derivatives corresponding to the formula M4-X1-Y1  
 XX are claimed (see features table). Sequences AAR66244-R66266 are specific  
 XX examples of such derivs. which have both improved fibrinolytic and  
 XX thrombin-inhibiting activities, compared to known plasminogen activators  
 XX or thrombin inhibitors. The proteins are useful as thrombolytic agents,  
 XX e.g. for treatment of arterial occlusions, deep vein thrombosis, cardiac  
 XX and cerebral infarction and pulmonary embolism. (Updated on 25-MAR-2003  
 XX to correct PN field.)  
 XX  
 XX SQ Sequence 390 AA;  
 XX  
 XX Query Match 100.0%; Score 44; DB 2; Length 390;  
 XX Best Local Similarity 100.0%; Pred. No. 98;  
 XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 XX QY 1 KPSSPPE 8  
 XX |||||  
 XX 90 KPSSPPE 97  
 XX  
 XX RESULT 24  
 XX AAR66247  
 XX ID AAR66247 standard; protein; 390 AA.  
 XX  
 XX AAR66247;  
 XX  
 XX 25-MAR-2003 (revised)  
 XX DT 17-AUG-1995 (first entry)  
 XX  
 XX Bifunctional urokinase variant M14.  
 XX  
 XX fibrinolysis; thrombin inhibition; thrombolytic; anti-thrombotic;  
 XX urokinase; variant; mutein.  
 XX  
 XX Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 XX FH 1..365  
 XX Region /label= M4  
 XX FT /note= "unglycosylated prourokinase(Ser47-Leu411)"  
 XX FT

```

FT Disulfide-bond 4. .85
FT Disulfide-bond 25. .67
FT Disulfide-bond 56. .80
FT Disulfide-bond 102. .233
FT Disulfide-bond 143. .159
FT Disulfide-bond 151. .222
FT Disulfide-bond 247. .316
FT Disulfide-bond 279. .295
FT Disulfide-bond 306. .334
FT Disulfide-bond 366. .371
FT Region /label= X1
FT Region /label= Y1
FT Region /label= Y1
FT DE4323754-C1.
FT 01-DEC-1994.
FT 15-JUL-1993; 93DE-04323754.
FT 15-JUL-1993; 93DE-04323754.
FT (CHEF ) GRUENENTHAL GMBH.
FT Steffens GJ, Wendt S, Schneider J, Heinzel-Wieland R;
FT Saunders DJ;
FT WPI; 1995-015191/03.
FT
FT New bifunctional urokinase derivs and related plasmids - with improved
FT fibrinolytic and thrombin inhibiting activities, for treating cardiac and
FT cerebral infarct, pulmonary embolism, etc.
FT
FT Example 1; Page 10 and Fig 1; 34pp; German.
FT
XX Bifunctional urokinase derivatives corresponding to the formula M4-X1-Y1
XX are claimed (see features table). Sequences AAR66244-R66266 are specific
XX examples of such derivs. which have both improved fibrinolytic and
XX thrombin-inhibiting activities, compared to known plasminogen activators
XX or thrombin inhibitors. The proteins are useful as thrombolytic agents,
XX e.g. for treatment of arterial occlusions, deep vein thrombosis, cardiac
XX and cerebral infarction and pulmonary embolism. (Updated on 25-MAR-2003
XX to correct PN field.)
XX
XX Sequence 390 AA;
XX
SQ
Query Match 100.0%; Score 44; DB 2; Length 390;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KPSSPPE 8
Db 90 KPSSPPE 97

```

RESULT 25  
AAR66260  
ID AAR66260 standard; protein; 392 AA.

XX AAR66260;  
XX  
XX 25-MAR-2003 (revised)  
XX 17-AUG-1995 (first entry)  
XX  
XX Bifunctional urokinase variant M27.  
XX  
XX fibrinolysis; thrombin inhibition; thrombolytic; anti-thrombotic;  
XX urokinase; variant; mutein.  
XX  
XX Synthetic.  
XX  
XX Key Location/Qualifiers  
XX Region 1. .365

```

FT /label= M4
FT /note= "unglycosylated prourokinase(Ser47-Leu411)"
FT Disulfide-bond 4. .85
FT Disulfide-bond 25. .67
FT Disulfide-bond 56. .80
FT Disulfide-bond 102. .233
FT Disulfide-bond 143. .159
FT Disulfide-bond 151. .222
FT Disulfide-bond 247. .316
FT Disulfide-bond 279. .295
FT Disulfide-bond 306. .334
FT Disulfide-bond 366. .371
FT Region /label= X1
FT Region /label= X1
FT Region /label= Y1
FT DE4323754-C1.
FT 01-DEC-1994.
FT 15-JUL-1993; 93DE-04323754.
FT 15-JUL-1993; 93DE-04323754.
FT (CHEF ) GRUENENTHAL GMBH.
FT Steffens GJ, Wendt S, Schneider J, Heinzel-Wieland R;
FT Saunders DJ;
FT WPI; 1995-015191/03.
FT
FT New bifunctional urokinase derivs and related plasmids - with improved
FT fibrinolytic and thrombin inhibiting activities, for treating cardiac and
FT cerebral infarct, pulmonary embolism, etc.
FT
FT Example 1; Page 11 and Fig 1; 34pp; German.
FT
XX Bifunctional urokinase derivatives corresponding to the formula M4-X1-Y1
XX are claimed (see features table). Sequences AAR66244-R66266 are specific
XX examples of such derivs. which have both improved fibrinolytic and
XX thrombin-inhibiting activities, compared to known plasminogen activators
XX or thrombin inhibitors. The proteins are useful as thrombolytic agents,
XX e.g. for treatment of arterial occlusions, deep vein thrombosis, cardiac
XX and cerebral infarction and pulmonary embolism. (Updated on 25-MAR-2003
XX to correct PN field.)
XX
XX Sequence 392 AA;
XX
SQ
Query Match 100.0%; Score 44; DB 2; Length 392;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KPSSPPE 8
Db 90 KPSSPPE 97

```

RESULT 26  
AAR66264  
ID AAR66264 standard; protein; 392 AA.

XX AAR66264;  
XX  
XX 25-MAR-2003 (revised)  
XX 17-AUG-1995 (first entry)  
XX  
XX Bifunctional urokinase variant M31.  
XX  
XX fibrinolysis; thrombin inhibition; thrombolytic; anti-thrombotic;  
XX urokinase; variant; mutein.  
XX  
XX Synthetic.  
XX  
XX Key Location/Qualifiers  
XX Region 1. .365

PH	Key	Location/Qualifiers
FT	Region	1..365
FT		/label= M4
FT		/note= "unglycosylated prourokinase (Ser47-Leu411)"
FT	Diulfide-bond	4..85
FT	Diulfide-bond	25..67
FT	Diulfide-bond	56..80
FT	Diulfide-bond	102..233
FT	Diulfide-bond	143..159
FT	Diulfide-bond	151..222
FT	Diulfide-bond	247..316
FT	Diulfide-bond	279..295
FT	Diulfide-bond	306..334
FT	Region	366..371
FT		/label= X1
FT	Region	372..392
FT		/label= Y1
PN	DE4323754-CL.	
XX		
XX	01-DEC-1994.	
XX		
XX	15-JUL-1993;	93DE-04323754.
XX		
XX	15-JUL-1993;	93DE-04323754.
XX		
XX	(CHEF ) GRUENENTHAL GMBH.	
XX		
XX	Steffens GJ, Wnendt S, Schneider J, Heinzel-Wieland R;	
PI	Saunders DJ;	
XX		
DR	WPI: 1995-015191/03.	
XX		
XX	New bifunctional urokinase derivs and related plasmids - with improved	
PT	fibrinolytic and thrombin inhibiting activities, for treating cardiac and	
PT	cerebral infarct, pulmonary embolism, etc.	
XX		
XX	Example 1, Page 11 and Fig 1, 34pp; German.	
XX		
XX	Bifunctional urokinase derivatives corresponding to the formula M4-X1-Y1	
CC	are claimed (see features table). Sequences AAR66244-R66266 are specific	
CC	examples of such derivs, which have both improved fibrinolytic and	
CC	thrombin-inhibiting activities, compared to known plasminogen activators	
CC	or thrombin inhibitors. The proteins are useful as thrombolytic agents,	
CC	e.g. for treatment of arterial occlusions, deep vein thrombosis, cardiac	
CC	and cerebral infarction and pulmonary embolism. (Updated on 25-MAR-2003	
CC	to correct PN field.)	
XX		
SO	Sequence 392 AA;	
	Query Match	100.0%; Score 44; DB 2; Length 392;
	Best Local Similarity	100.0%; Pred. No. 98;
	Matches	8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 KPSSPPEE 8	
	90 KPSSPPEE 97	
DB		
	RESULT 27	
ID	AAR66255	
XX	AAR66255 standard; protein; 392 AA.	
XX		
AC	AAR66255;	
XX		
DT	25-MAR-2003 (revised)	
DT	17-AUG-1995 (first entry)	
XX		
XX	Bifunctional urokinase variant M22.	
DE		
XX	fibrinolysis; thrombin inhibition; thrombolytic; anti-thrombotic;	
KW	urokinase; variant; mutein.	
XX		

[illegible]

```

KW urokinase; variant; mutein.
XX
XX Synthetic.
OS
XX
XX
FH Key Location/Qualifiers
FT Region 1..365
FT /label= M4
FT /note= "unglycosylated prourokinase(Ser47-Leu411)"
FT Disulfide-bond 4..85
FT Disulfide-bond 25..67
FT Disulfide-bond 56..80
FT Disulfide-bond 102..233
FT Disulfide-bond 143..159
FT Disulfide-bond 151..222
FT Disulfide-bond 247..316
FT Disulfide-bond 279..295
FT Disulfide-bond 306..334
FT Disulfide-bond 366..371
FT Region
FT /label= X1
FT /label= X1
FT /label= Y1
FT /label= Y1
PN DE4323754-Cl.
PD 01-DEC-1994.
XX
XX 15-JUL-1993; 93DE-04323754.
PF
XX
XX 15-JUL-1993; 93DE-04323754.
PR
XX
XX (CHEF ) GRUENENTHAL GMBH.
XX
XX Steffens GJ, Wnendt S, Schneider J, Heinzel-Wieland R;
XX Saunders DJ;
XX WPI; 1995-015191/03.
DR
XX
XX New bifunctional urokinase derivs and related plasmids - with improved
XX fibrinolytic and thrombin inhibiting activities, for treating cardiac and
XX cerebral infarct, pulmonary embolism, etc.
XX
XX Example 1; Page 11 and Fig 1; 34pp; German.
XX
XX Bifunctional urokinase derivatives corresponding to the formula M4-X1-Y1
XX are claimed (see features table). Sequences AAR66244-R66266 are specific
XX examples of such derivs. which have both improved fibrinolytic and
XX thrombin-inhibiting activities, compared to known plasminogen activators
XX or thrombin inhibitors. The proteins are useful as thrombolytic agents,
XX e.g. for treatment of arterial occlusions, deep vein thrombosis, cardiac
XX and cerebral infarction and pulmonary embolism. (Updated on 25-MAR-2003
XX to correct PN field.)
XX
XX Sequence 392 AA;
SQ
Query Match 100.0%; Score 44; DB 2; Length 392;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KPSSPPE 8
DB 90 KPSSPPE 97

```

```

RESULT 29
AAR66261
ID AAR66261 standard; protein; 392 AA.
XX
XX AAR66261;
AC
XX
XX 25-MAR-2003 (revised)
DT 17-AUG-1995 (first entry)
XX
XX Bifunctional urokinase variant M28.
DE

```

```

XX
XX fibrinolysis; thrombin inhibition; chrombolytic; anti-chrombolic;
KW urokinase; variant; mutein.
XX
XX Synthetic.
OS
XX
XX
FH Key Location/Qualifiers
FT Region 1..365
FT /label= M4
FT /note= "unglycosylated prourokinase(Ser47-Leu411)"
FT Disulfide-bond 4..85
FT Disulfide-bond 25..67
FT Disulfide-bond 56..80
FT Disulfide-bond 102..233
FT Disulfide-bond 143..159
FT Disulfide-bond 151..222
FT Disulfide-bond 247..316
FT Disulfide-bond 279..295
FT Disulfide-bond 306..334
FT Disulfide-bond 366..371
FT Region
FT /label= X1
FT /label= X1
FT /label= Y1
FT /label= Y1
PN DE4323754-Cl.
PD 01-DEC-1994.
XX
XX 15-JUL-1993; 93DE-04323754.
PF
XX
XX 15-JUL-1993; 93DE-04323754.
PR
XX
XX (CHEF ) GRUENENTHAL GMBH.
XX
XX Steffens GJ, Wnendt S, Schneider J, Heinzel-Wieland R;
XX Saunders DJ;
XX WPI; 1995-015191/03.
DR
XX
XX New bifunctional urokinase derivs and related plasmids - with improved
XX fibrinolytic and thrombin inhibiting activities, for treating cardiac and
XX cerebral infarct, pulmonary embolism, etc.
XX
XX Example 1; Page 11 and Fig 1; 34pp; German.
XX
XX Bifunctional urokinase derivatives corresponding to the formula M4-X1-Y1
XX are claimed (see features table). Sequences AAR66244-R66266 are specific
XX examples of such derivs. which have both improved fibrinolytic and
XX thrombin-inhibiting activities, compared to known plasminogen activators
XX or thrombin inhibitors. The proteins are useful as thrombolytic agents,
XX e.g. for treatment of arterial occlusions, deep vein thrombosis, cardiac
XX and cerebral infarction and pulmonary embolism. (Updated on 25-MAR-2003
XX to correct PN field.)
XX
XX Sequence 392 AA;
SQ
Query Match 100.0%; Score 44; DB 2; Length 392;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KPSSPPE 8
DB 90 KPSSPPE 97

```

```

RESULT 30
AAR66258
ID AAR66258 standard; protein; 392 AA.
XX
XX AAR66258;
AC
XX
XX 25-MAR-2003 (revised)
DT 17-AUG-1995 (first entry)
XX
XX

```

```

XX DE Bifunctional urokinase variant M25.
XX KM fibrinolysis; thrombin inhibition; thrombolytic; anti-thrombotic;
XX KM urokinase; variant; mutein.
XX OS Synthetic.
XX FH Key
XX FT Region
XX FT Location/Qualifiers
XX FT /label= M4
XX FT /note= "unglycosylated prourokinase (Ser47-Leu411)"
XX FT Disulfide-bond 4. .85
XX FT Disulfide-bond 25. .67
XX FT Disulfide-bond 56. .80
XX FT Disulfide-bond 102. .233
XX FT Disulfide-bond 143. .159
XX FT Disulfide-bond 151. .222
XX FT Disulfide-bond 247. .316
XX FT Disulfide-bond 279. .295
XX FT Disulfide-bond 306. .334
XX FT Disulfide-bond 366. .371
XX FT Region
XX FT /label= X1
XX FT /label= X1
XX FT Region
XX FT /label= Y1
XX PN DE4323754-Cl.
XX PD 01-DEC-1994.
XX PF 15-JUL-1993; 93DE-04323754.
XX PR 15-JUL-1993; 93DE-04323754.
XX PA (CHEF ) GRUENENTHAL GMBH.
XX PI Steffens GJ, Wnendt S, Schneider J, Heinzel-Wieland R;
XX PI Saunders DJ;
XX WPI; 1995-015191/03.
XX DR
XX XX New bifunctional urokinase derivs and related plasmids - with improved
XX PT fibrinolytic and thrombin inhibiting activities, for treating cardiac and
XX PT cerebral infarct, pulmonary embolism, etc.
XX PS Example 1; Page 11 and Fig 1; 34pp; German.
XX PS Bifunctional urokinase derivatives corresponding to the formula M4-X1-Y1
XX CC are claimed (see features table). Sequences AAR66244-R66266 are specific
XX CC examples of such derivs. which have both improved fibrinolytic and
XX CC thrombin-inhibiting activities, compared to known plasminogen activators
XX CC or thrombin inhibitors. The proteins are useful as thrombolytic agents,
XX CC e.g. for treatment of arterial occlusions, deep vein thrombosis, cardiac
XX CC and cerebral infarction and pulmonary embolism. (Updated on 25-MAR-2003
XX CC to correct PN field.)
XX SQ Sequence 392 AA;
XX
XX Query Match 100.0%; Score 44; DB 2; Length 392;
XX Best Local Similarity 100.0%; Pred. No. 98;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KPSSPPE 8
Db 90 KPSSPPE 97

```

```

DT 25-MAR-2003 (revised)
DT 17-AUG-1995 (first entry)
XX DE Bifunctional urokinase variant M30.
XX KM fibrinolysis; thrombin inhibition; thrombolytic; anti-thrombotic;
XX KM urokinase; variant; mutein.
XX OS Synthetic.
XX FH Key
XX FT Region
XX FT Location/Qualifiers
XX FT /label= M4
XX FT /note= "unglycosylated prourokinase (Ser47-Leu411)"
XX FT Disulfide-bond 4. .85
XX FT Disulfide-bond 25. .67
XX FT Disulfide-bond 56. .80
XX FT Disulfide-bond 102. .233
XX FT Disulfide-bond 143. .159
XX FT Disulfide-bond 151. .222
XX FT Disulfide-bond 247. .316
XX FT Disulfide-bond 279. .295
XX FT Disulfide-bond 306. .334
XX FT Disulfide-bond 366. .371
XX FT Region
XX FT /label= X1
XX FT /label= X1
XX FT Region
XX FT /label= Y1
XX PN DE4323754-Cl.
XX PD 01-DEC-1994.
XX PF 15-JUL-1993; 93DE-04323754.
XX PR 15-JUL-1993; 93DE-04323754.
XX PA (CHEF ) GRUENENTHAL GMBH.
XX PI Steffens GJ, Wnendt S, Schneider J, Heinzel-Wieland R;
XX PI Saunders DJ;
XX WPI; 1995-015191/03.
XX DR
XX XX New bifunctional urokinase derivs and related plasmids - with improved
XX PT fibrinolytic and thrombin inhibiting activities, for treating cardiac and
XX PT cerebral infarct, pulmonary embolism, etc.
XX PS Example 1; Page 11 and Fig 1; 34pp; German.
XX PS Bifunctional urokinase derivatives corresponding to the formula M4-X1-Y1
XX CC are claimed (see features table). Sequences AAR66244-R66266 are specific
XX CC examples of such derivs. which have both improved fibrinolytic and
XX CC thrombin-inhibiting activities, compared to known plasminogen activators
XX CC or thrombin inhibitors. The proteins are useful as thrombolytic agents,
XX CC e.g. for treatment of arterial occlusions, deep vein thrombosis, cardiac
XX CC and cerebral infarction and pulmonary embolism. (Updated on 25-MAR-2003
XX CC to correct PN field.)
XX SQ Sequence 392 AA;
XX
XX Query Match 100.0%; Score 44; DB 2; Length 392;
XX Best Local Similarity 100.0%; Pred. No. 98;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KPSSPPE 8
Db 90 KPSSPPE 97

```

RESULT 31  
 ID AAR66263 standard; protein; 392 AA.  
 XX AAR66263;  
 XX

RESULT 32  
 ID AAR66254 standard; protein; 392 AA.  
 XX AAR66254;  
 XX

```

AC AAR66254;
XX
XX 25-MAR-2003 (revised)
DT 17-AUG-1995 (first entry)
XX
XX Bifunctional urokinase variant M21.
XX
XX fibrinolysis; thrombin inhibition; thrombolytic; anti-thrombotic;
KM urokinase; variant; mutein.
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FH 1.365
FT Region /label= M4
FT /note= "unglycosylated prourokinase(Ser47-Leu411)"
FT Disulfide-bond 4..85
FT Disulfide-bond 25..67
FT Disulfide-bond 56..80
FT Disulfide-bond 102..233
FT Disulfide-bond 143..159
FT Disulfide-bond 151..222
FT Disulfide-bond 247..316
FT Disulfide-bond 279..295
FT Disulfide-bond 306..334
FT Disulfide-bond 366..371
FT Region /label= X1
FT 372..392
FT Region /label= Y1
XX
XX DE4323754-C1.
XX
XX 01-DEC-1994.
XX
XX 15-JUL-1993; 93DE-04323754.
XX
XX 15-JUL-1993; 93DE-04323754.
XX
XX (CHEF ) GRUENENTHAL GMBH.
XX
XX Steffens GJ, Wnendt S, Schneider J, Heinzel-Wieland R;
PI Saunders DJ;
XX
XX WPI; 1995-015191/03.
XX
XX New bifunctional urokinase derivs and related plasmids - with improved
PT fibrinolytic and thrombin inhibiting activities, for treating cardiac and
PT cerebral infarct, pulmonary embolism, etc.
XX
XX Example 1; Page 10 and Fig 1; 34pp; German.
XX
XX Bifunctional urokinase derivatives corresponding to the formula M4-X1-Y1
CC are claimed (see features table). Sequences AAR66244-R66266 are specific
CC examples of such derivs. which have both improved fibrinolytic and
CC thrombin-inhibiting activities, compared to known plasminogen activators
CC or thrombin inhibitors. The proteins are useful as thrombolytic agents,
CC e.g. for treatment of arterial occlusions, deep vein thrombosis, cardiac
CC and cerebral infarction and pulmonary embolism. (Updated on 25-MAR-2003
CC to correct PN field.)
XX
XX Sequence 392 AA;
SQ
Query Match 100.0%; Score 44; DB 2; Length 392;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KPSSPPE 8
Db 90 KPSSPPE 97

```

RESULT 33  
AAR66256

```

ID AAR66256 standard; protein; 392 AA.
XX
XX AAR66256;
AC
XX 25-MAR-2003 (revised)
DT 17-AUG-1995 (first entry)
XX
XX Bifunctional urokinase variant M23.
XX
XX fibrinolysis; thrombin inhibition; thrombolytic; anti-thrombotic;
KM urokinase; variant; mutein.
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FH 1.365
FT Region /label= M4
FT /note= "unglycosylated prourokinase(Ser47-Leu411)"
FT Disulfide-bond 4..85
FT Disulfide-bond 25..67
FT Disulfide-bond 56..80
FT Disulfide-bond 102..233
FT Disulfide-bond 143..159
FT Disulfide-bond 151..222
FT Disulfide-bond 247..316
FT Disulfide-bond 279..295
FT Disulfide-bond 306..334
FT Disulfide-bond 366..371
FT Region /label= X1
FT 372..392
FT Region /label= Y1
XX
XX DE4323754-C1.
XX
XX 01-DEC-1994.
XX
XX 15-JUL-1993; 93DE-04323754.
XX
XX 15-JUL-1993; 93DE-04323754.
XX
XX (CHEF ) GRUENENTHAL GMBH.
XX
XX Steffens GJ, Wnendt S, Schneider J, Heinzel-Wieland R;
PI Saunders DJ;
XX
XX WPI; 1995-015191/03.
XX
XX New bifunctional urokinase derivs and related plasmids - with improved
PT fibrinolytic and thrombin inhibiting activities, for treating cardiac and
PT cerebral infarct, pulmonary embolism, etc.
XX
XX Example 1; Page 10 and Fig 1; 34pp; German.
XX
XX Bifunctional urokinase derivatives corresponding to the formula M4-X1-Y1
CC are claimed (see features table). Sequences AAR66244-R66266 are specific
CC examples of such derivs. which have both improved fibrinolytic and
CC thrombin-inhibiting activities, compared to known plasminogen activators
CC or thrombin inhibitors. The proteins are useful as thrombolytic agents,
CC e.g. for treatment of arterial occlusions, deep vein thrombosis, cardiac
CC and cerebral infarction and pulmonary embolism. (Updated on 25-MAR-2003
CC to correct PN field.)
XX
XX Sequence 392 AA;
SQ
Query Match 100.0%; Score 44; DB 2; Length 392;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KPSSPPE 8
Db 90 KPSSPPE 97

```



```

RESULT 34
AAR66257
ID AAR66257 standard; protein; 392 AA.
XX
AC AAR66257;
XX
DT 25-MAR-2003 (revised)
DT 17-AUG-1995 (first entry)
XX
DE Bifunctional urokinase variant M24.
XX
KW fibrinolysis; thrombin inhibition; chromolytic; anti-chromobotic;
KW urokinase; variant; mutein.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FH Region 1..365
FT /label= M4
FT /note= "unglycosylated prourokinase(Ser47-Leu411)"
FT Disulfide-bond 4..85
FT Disulfide-bond 25..67
FT Disulfide-bond 56..80
FT Disulfide-bond 102..233
FT Disulfide-bond 143..159
FT Disulfide-bond 151..222
FT Disulfide-bond 247..316
FT Disulfide-bond 279..295
FT Disulfide-bond 306..334
FT Disulfide-bond 366..371
FT Region /label= X1
FT /label= Y1
FT Region /label= Y1
XX
PN DE4323754-Cl.
XX
PD 01-DEC-1994.
XX
PF 15-JUL-1993; 93DE-04323754.
XX
PR 15-JUL-1993; 93DE-04323754.
XX
PA (CHER ) GRUENTHAL GMBH.
XX
PI Steffens GJ, Wnendt S, Schneider J, Heinzel-Wieland R;
PI Saunders DJ;
XX
DR WPI; 1995-015191/03.
XX
PT New bifunctional urokinase derivs and related plasmids - with improved
PT fibrinolytic and thrombin inhibiting activities, for treating cardiac and
PT cerebral infarct, pulmonary embolism, etc.
XX
PS Example 1; Page 11 and Fig 1; 34pp; German.
XX
CC Bifunctional urokinase derivatives corresponding to the formula M4-X1-Y1
CC are claimed (see features table). Sequences AAR66244-R66266 are specific
CC examples of such derivs. which have both improved fibrinolytic and
CC thrombin-inhibiting activities, compared to known plasminogen activators
CC or thrombin inhibitors. The proteins are useful as thrombolytic agents,
CC e.g. for treatment of arterial occlusions, deep vein thrombosis, cardiac
CC and cerebral infarction and pulmonary embolism. (Updated on 25-MAR-2003
CC to correct PN field.)
XX
SQ Sequence 392 AA;
XX
Query Match 100.0%; Score 44; DB 2; Length 392;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KPSSPPE 8
DB 90 KPSSPPE 97

```

```

RESULT 35
AAR66251
ID AAR66251 standard; protein; 393 AA.
XX
AC AAR66251;
XX
DT 25-MAR-2003 (revised)
DT 17-AUG-1995 (first entry)
XX
DE Bifunctional urokinase variant M18.
XX
KW fibrinolysis; thrombin inhibition; chromolytic; anti-chromobotic;
KW urokinase; variant; mutein.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FH Region 1..365
FT /label= M4
FT /note= "unglycosylated prourokinase(Ser47-Leu411)"
FT Disulfide-bond 4..85
FT Disulfide-bond 25..67
FT Disulfide-bond 56..80
FT Disulfide-bond 102..233
FT Disulfide-bond 143..159
FT Disulfide-bond 151..222
FT Disulfide-bond 247..316
FT Disulfide-bond 279..295
FT Disulfide-bond 306..334
FT Disulfide-bond 366..372
FT Region /label= X1
FT /label= Y1
FT Region /label= Y1
XX
PN DE4323754-Cl.
XX
PD 01-DEC-1994.
XX
PF 15-JUL-1993; 93DE-04323754.
XX
PR 15-JUL-1993; 93DE-04323754.
XX
PA (CHER ) GRUENTHAL GMBH.
XX
PI Steffens GJ, Wnendt S, Schneider J, Heinzel-Wieland R;
PI Saunders DJ;
XX
DR WPI; 1995-015191/03.
XX
PT New bifunctional urokinase derivs and related plasmids - with improved
PT fibrinolytic and thrombin inhibiting activities, for treating cardiac and
PT cerebral infarct, pulmonary embolism, etc.
XX
PS Example 1; Page 10 and Fig 1; 34pp; German.
XX
CC Bifunctional urokinase derivatives corresponding to the formula M4-X1-Y1
CC are claimed (see features table). Sequences AAR66244-R66266 are specific
CC examples of such derivs. which have both improved fibrinolytic and
CC thrombin-inhibiting activities, compared to known plasminogen activators
CC or thrombin inhibitors. The proteins are useful as thrombolytic agents,
CC e.g. for treatment of arterial occlusions, deep vein thrombosis, cardiac
CC and cerebral infarction and pulmonary embolism. (Updated on 25-MAR-2003
CC to correct PN field.)
XX
SQ Sequence 393 AA;
XX
Query Match 100.0%; Score 44; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KPSSPPE 8

```

Db 90 KPSSPPE 97

RESULT 36

AA66253 standard; protein; 393 AA.

AC AA66253;

DT 25-MAR-2003 (revised)

DT 17-AUG-1995 (first entry)

DE Bifunctional urokinase variant M20.

KM fibrinolysis; thrombin inhibition; thrombolytic; anti-thrombotic; urokinase; variant; mutein.

OS Synthetic.

Key Location/Qualifiers  
FH 1..365  
FT /label= M4

FT /note= "unglycosylated prourokinase(Ser47-Leu411)"

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

Qy 1 KPSSPPE 8  
90 KPSSPPE 97

RESULT 37

AA66249 standard; protein; 393 AA.

AC AA66249;

DT 25-MAR-2003 (revised)

DT 17-AUG-1995 (first entry)

DE Bifunctional urokinase variant M16.

KM fibrinolysis; thrombin inhibition; thrombolytic; anti-thrombotic; urokinase; variant; mutein.

OS Synthetic.

Key Location/Qualifiers  
FH 1..365  
FT /label= M4

FT /note= "unglycosylated prourokinase(Ser47-Leu411)"

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

FT Disulfide-bond

Query Match 100.0%; Score 44; DB 2; Length 393;

Best Local Similarity 100.0%; Pred. No. 99; Mismatches 0; Indels 0; Gaps 0;

Query Match 100.0%; Score 44; DB 2; Length 393;

Best Local Similarity 100.0%; Pred. No. 99; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 100.0%; Pred. No. 99;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8  
| | | | | | | |  
DB 90 KPSSPPEE 97

## RESULT 38

AA66252  
ID AA66252 standard; protein; 393 AA.

AC AA66252;

DT 25-MAR-2003 (revised)

DT 17-AUG-1995 (first entry)

DE Bifunctional urokinase variant M19.

KW fibrinolysis; thrombin inhibition; thrombolytic; anti-thrombotic;  
urokinase; variant; mutein.

OS Synthetic.

Key Location/Qualifiers  
Region 1..365  
/label= M4

FT Disulfide-bond 4..85  
/note="unglycosylated prourokinase(Ser47-Leu411)"

FT Disulfide-bond 25..67

FT Disulfide-bond 56..80

FT Disulfide-bond 102..233

FT Disulfide-bond 143..159

FT Disulfide-bond 151..222

FT Disulfide-bond 247..316

FT Disulfide-bond 279..295

FT Disulfide-bond 306..334

FT Disulfide-bond 366..372

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

FT Disulfide-bond 373..393

Query Match 100.0%; Score 44; DB 2; Length 393;

Best Local Similarity 100.0%; Pred. No. 99;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8  
| | | | | | | |  
DB 90 KPSSPPEE 97

## RESULT 39

AA66244  
ID AA66244 standard; protein; 393 AA.

AC AA66244;

DT 25-MAR-2003 (revised)

DT 22-AUG-1995 (first entry)

DE Bifunctional urokinase variant M11.

KW fibrinolysis; thrombin inhibition; thrombolytic; anti-thrombotic;  
urokinase; variant; mutein.

OS Synthetic.

Key Location/Qualifiers  
Region 1..365  
/label= M4

FT Disulfide-bond 4..85  
/note="unglycosylated prourokinase(Ser47-Leu411)"

FT Disulfide-bond 25..67

FT Disulfide-bond 56..80

FT Disulfide-bond 102..233

FT Disulfide-bond 143..159

FT Disulfide-bond 151..222

FT Disulfide-bond 247..316

FT Disulfide-bond 279..295

FT Disulfide-bond 306..334

FT Disulfide-bond 366..374

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

FT Disulfide-bond 375..393

New bifunctional urokinase derivs and related plasmids - with improved  
fibrinolytic and thrombin inhibiting activities, for treating cardiac and  
cerebral infarct, pulmonary embolism, etc.

WPI, 1995-015191/03.

Example 1; Page 10 and Fig 1; 34pp; German.

Bifunctional urokinase derivatives corresponding to the formula M4-X1-Y1

are claimed (see features table). Sequences AA66244-R66266 are specific

examples of such derivs. which have both improved fibrinolytic and

thrombin-inhibiting activities, compared to known plasminogen activators

or thrombin inhibitors. The proteins are useful as thrombolytic agents,

e.g. for treatment of arterial occlusions, deep vein thrombosis, cardiac

and cerebral infarction and pulmonary embolism. (Updated on 25-MAR-2003

to correct PN field.)

Sequence 393 AA;

XX SQ Sequence 393 AA;  
 Query Match 100.0%; Score 44; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 99;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8  
 |||||  
 90 KPSSPPEE 97

Db 90 KPSSPPEE 97

RESULT 40  
 AAR6250  
 ID AAR6250 standard; protein; 393 AA.  
 XX AAR6250;  
 AC AAR6250;  
 DT 25-MAR-2003 (revised)  
 DT 17-AUG-1995 (first entry)  
 XX  
 DE Bifunctional urokinase variant M17.  
 XX  
 KM fibrinolysis; thrombin inhibition; thrombolytic; anti-thrombotic;  
 KM urokinase; variant; muten.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 1..365  
 FT /label= M4  
 FT /note="unglycosylated prourokinase(Ser47-Leu411)"  
 FT Disulfide-bond 4..85  
 FT Disulfide-bond 25..67  
 FT Disulfide-bond 56..80  
 FT Disulfide-bond 102..233  
 FT Disulfide-bond 143..159  
 FT Disulfide-bond 151..222  
 FT Disulfide-bond 247..316  
 FT Disulfide-bond 279..295  
 FT Disulfide-bond 306..334  
 FT Region 366..372  
 FT /label= X1  
 FT Region 373..393  
 FT /label= Y1  
 XX  
 PN DE4323754-C1.  
 PD 01-DEC-1994.  
 PF 15-JUL-1993; 93DE-04323754.  
 PR 15-JUL-1993; 93DE-04323754.  
 XX  
 PA (CHEF ) GRUENENTHAL GMBH.  
 PI Steffens GJ, Wendt S, Schneider J, Heinzel-Wieland R;  
 PI Saunders DJ;  
 XX  
 DR WPI; 1995-015191/03.  
 XX  
 PT New bifunctional urokinase derives and related plasmids - with improved  
 PT fibrinolytic and thrombin inhibiting activities, for treating cardiac and  
 PT cerebral infarct, pulmonary embolism, etc.  
 XX  
 PS Example 1; Page 10 and Fig 1; 34pp; German.  
 XX  
 CC Bifunctional urokinase derivatives corresponding to the formula M4-X1-Y1  
 CC are claimed (see features table). Sequences AAR6250-R6266 are specific  
 CC examples of such derivs. which have both improved fibrinolytic and  
 CC thrombin-inhibiting activities, compared to known plasminogen activators  
 CC or thrombin inhibitors. The proteins are useful as thrombolytic agents,  
 CC e.g. for treatment of arterial occlusions, deep vein thrombosis, cardiac

CC and cerebral infarction and pulmonary embolism. (Updated on 25-MAR-2003  
 CC to correct PN field.)  
 XX  
 XX SQ Sequence 393 AA;  
 Query Match 100.0%; Score 44; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 99;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8  
 |||||  
 90 KPSSPPEE 97

Db 90 KPSSPPEE 97

RESULT 41  
 AAR9597  
 ID AAR9597 standard; protein; 393 AA.  
 XX AAR9597;  
 AC AAR9597;  
 DT 05-DEC-1996 (first entry)  
 DT  
 XX  
 DE Chimeric protein M38 encoded by pSEI.  
 XX  
 KM Thrombin; inhibition; thrombus; thrombolysis; chimeric protein;  
 KM plasminogen activating sequence; fibrinolysis; infarction;  
 KM angina pectoris; deep vein thrombosis.  
 XX  
 OS Synthetic.  
 XX  
 PN EP714982-A2.  
 XX  
 PD 05-JUN-1996.  
 PF 16-NOV-1995; 95EP-00118050.  
 XX  
 PR 30-NOV-1994; 94DE-04442665.  
 XX  
 PA (CHEF ) GRUENENTHAL GMBH.  
 PI Wendt S, Steffens GJ, Janocha E, Heinzel-Wieland R;  
 XX  
 DR WPI; 1996-269715/28.  
 XX  
 PT Chimeric protein contg. plasminogen activating sequence and thrombin-  
 PT inhibiting sequence - useful as thrombus-specific thrombolytic agent with  
 PT rapid action.  
 XX  
 PS Example 1; Page 21-22; 37pp; German.  
 XX  
 CC Example 1 describes the prodn. of plasmids pSEI and pS89 contg. a DNA  
 CC encoding a chimeric protein with fibrinolytic and thrombin-inhibiting  
 CC properties. pSEI encodes the protein given in AAR9597 and pS89 encodes  
 CC the protein given in AAR9596  
 XX  
 SQ Sequence 393 AA;  
 Query Match 100.0%; Score 44; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 99;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8  
 |||||  
 91 KPSSPPEE 98

Db 91 KPSSPPEE 98

RESULT 42  
 AAR9596  
 ID AAR9596 standard; protein; 393 AA.  
 XX AAR9596;  
 AC AAR9596;  
 DT 05-DEC-1996 (first entry)

XX Chimeric protein M37 encoded by pSE9.  
 DE Thrombin; inhibition; thrombus; thrombolysis; chimeric protein;  
 XX Thrombin; inhibition; thrombus; thrombolysis; chimeric protein;  
 KW plasminogen activating sequence; fibrinolysis; infarction;  
 KW angina pectoris; deep vein thrombosis.  
 XX Synthetic.  
 OS  
 XX EP714982-A2.  
 PN  
 XX 05-JUN-1996.  
 PD  
 XX 16-NOV-1995; 95EP-00118050.  
 PF  
 XX 30-NOV-1994; 94DE-04442665.  
 PR  
 XX (CHEF ) GRUENENTHAL GMBH.  
 PA  
 XX Wnendt S, Steffens GJ, Janocha E, Heinzel-Wieland R;  
 PI WPI; 1996-269715/28.  
 DR  
 XX Chimeric protein contg. plasminogen activating sequence and thrombin-  
 PT inhibiting sequence - useful as thrombus-specific thrombolytic agent with  
 PT rapid action.  
 PS Example 1; Page 19-20; 37p; German.  
 XX Example 1 describes the prodn. of plasmin pSE1 and pSE9 contg. a DNA  
 CC encoding a chimeric protein with fibrinolytic and thrombin-inhibiting  
 CC properties. pSE1 encodes the protein given in AAR99597 and pSE9 encodes  
 CC the protein given in AAR99596  
 SQ Sequence 393 AA;  
 Query Match 100.0%; Score 44; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 99;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KPSSPPEE 8  
 DB 91 KPSSPPEE 98  
 RESULT 43  
 AAR99885  
 ID AAR99885 standard; peptide; 393 AA.  
 XX  
 AC AAR99885;  
 XX  
 XX 27-JUN-1997 (first entry)  
 DT  
 XX M36: fibrinolytic and anticoagulant activity contg. protein.  
 DE  
 XX Thrombin; factor Xa; anticoagulant; thrombolytic; plasminogen; urokinase;  
 KW activator; streptokinase; staphylokinase; APSAC;  
 KW antisolated plasminogen streptokinase activator complex; hirudin;  
 KW hirudin; antistatin; pM127; pMS1; pSE8; pMS56.  
 XX  
 OS Synthetic.  
 XX  
 XX EP712934-A2.  
 PN  
 XX 22-MAY-1996.  
 PD  
 XX 03-NOV-1995; 95EP-00117316.  
 PF  
 XX 17-NOV-1994; 94DE-04440892.  
 PR  
 XX (CHEF ) GRUENENTHAL GMBH.  
 PA  
 XX Wnendt S, Heinzel-Wieland R, Steffens GJ,  
 PI

XX WPI; 1996-240720/25.  
 DR  
 XX Proteins with fibrinolytic and anticoagulant activity - useful as  
 PT thrombolytic agents.  
 XX  
 XX Disclosure; Fig 18; 59pp; German.  
 PS  
 XX New peptide (I) with fibrinolytic and anticoagulant activity comprise a  
 CC plasminogen-activating amino acid sequence (A) fused at the N- and/or C-  
 CC terminus to a thrombin and/or factor Xa inhibiting amino acid sequence  
 CC (B). Excluded from the claims are (I) where (A) is Ser47 to Leu411 of  
 CC unglycosylated urokinase linked at the C-terminus to sequences (i) to  
 CC (iii): T1-RE-T2-GGQNDPFEIPEBYL-T3 (i) T1-REFLRNPDKYEPWEDKKE (ii)  
 CC T1-RPSSRPEREIDEERK (iii) where T1 = P or V; T2 = L or a bond; T3 = Q or  
 CC OH. (A) is pref. (pro)urokinase, tissue plasminogen activator (tPA), bat-  
 CC PA (all opt. modified by deletion, substitution, insertion and/or addn.);  
 CC streptokinase; staphylokinase; and/or APSAC (antisolated plasminogen  
 CC streptokinase activator complex), esp. prourokinase (411 amino acids) or  
 CC its Ser47 to Leu411 or Ser138 to Leu411 fragments, or t-PA (527 amino  
 CC acids) or its Ser99Arg to 527Pro or 174Ser to 527Pro fragments. (B) has  
 CC hirudin or hirulin activities; or is derived from the human thrombin  
 CC receptor, antistatin and/or the tick anticoagulant peptide. Most pref.  
 CC are the 65 amino acid hirudin sequence or one of the six sequences given  
 CC in AAR99879 to AAR99884. Plasmids pM127 (MS1), pMS1 (MS112), pSE8 (M36)  
 CC and pMS56 (M43) contain the sequences encoding AAR99885 to AAR99888,  
 CC respectively. The products were tested in human citrated plasma (5 microg  
 CC in 200 microl 1:10 diluted plasma). The thrombin time was then 1.2, 3,  
 CC 2.8 and 1.2 times greater, respectively, than in the absence of the  
 CC product  
 SQ Sequence 393 AA;  
 Query Match 100.0%; Score 44; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 99;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KPSSPPEE 8  
 DB 91 KPSSPPEE 98  
 RESULT 44  
 AAR47902  
 ID AAR47902 standard; protein; 395 AA.  
 XX  
 AC AAR47902;  
 XX  
 XX 13-JUL-1994 (first entry)  
 DT  
 XX Pro-urokinase derivative.  
 DE  
 XX Pro-urokinase derivative.  
 XX  
 XX Pro-urokinase; half-life; thrombolytic; thrombosis; fibrinolytic; factor.  
 KW  
 XX Homo sapiens.  
 OS  
 XX JP05336965-A.  
 PN  
 XX 21-DEC-1993.  
 PD  
 XX 17-OCT-1991; 91JP-00269615.  
 PF  
 XX 17-OCT-1991; 91JP-00269615.  
 PR  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 PA  
 XX WPI; 1994-030907/04.  
 DR  
 XX N-PADB; AAO55771.  
 PT  
 XX Novel human pro-urokinase derivate, having long half-life - with high  
 PT thrombolytic activity, useful for treatment of thrombosis.  
 XX  
 PS Disclosure; Page 14; 29pp; Japanese.

XX Sequences (AAQ55771-72) are pro-urokinase derivatives. The products have  
CC an inserted sugar moiety having an amino acid substituted, depleted or  
CC inserted variant around the thrombin cleavage site. They also have a long  
CC half-life allowing them to be used in the treatment of thrombosis  
XX  
SQ Sequence 395 AA;

Query Match 100.0%; Score 44; DB 2; Length 395;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
Db 120 KPSSPPE 127

## RESULT 45

AA66265  
ID AAR66265 standard; protein; 395 AA.

AC AAR66265;

DT 25-MAR-2003 (revised)  
DT 17-AUG-1995 (first entry)

DE Bifunctional urokinase variant M32.

XX  
KM fibrinolysis; thrombin inhibition; thrombolytic; anti-thrombotic;  
KM urokinase; variant; mutein.

OS Synthetic.

XX  
FH Key Location/Qualifiers  
FT Region 1..365  
/label= M4  
/note= "unglycosylated prourokinase (Ser47-Leu411)"

FT Disulfide-bond 4..85

FT Disulfide-bond 25..67

FT Disulfide-bond 56..80

FT Disulfide-bond 102..233

FT Disulfide-bond 143..159

FT Disulfide-bond 151..222

FT Disulfide-bond 247..316

FT Disulfide-bond 279..295

FT Disulfide-bond 306..334

FT Disulfide-bond 366..371

FT Region /label= X1

FT Region 372..395

FT Region /label= Y1

PN DE4323754-C1.

PD 01-DEC-1994.

PF 15-JUL-1993; 93DE-04323754.

PR 15-JUL-1993; 93DE-04323754.

XX (CHEF ) GRUENENTHAL GMBH.

XX Steffens GJ, Wendt S, Schneider J, Heinzel-Wieland R;  
XX Saunders DJ;  
XX WPI; 1995-015191/03.

XX New bifunctional urokinase derivs and related plasmids - with improved  
XX fibrinolytic and thrombin inhibiting activities, for treating cardiac and  
XX cerebral infarct, pulmonary embolism, etc.  
XX Example 1; Page 11 and Fig 1; 34pp; German.  
XX Bifunctional urokinase derivatives corresponding to the formula M4-X1-Y1

CC are claimed (see features table). Sequences AAR66264-R66266 are specific  
CC examples of such derivs. which have both improved fibrinolytic and  
CC thrombin-inhibiting activities, compared to known plasminogen activators  
CC or thrombin inhibitors. The proteins are useful as thrombolytic agents,  
CC e.g. for treatment of arterial occlusions, deep vein thrombosis, cardiac  
CC and cerebral infarction and pulmonary embolism. (Updated on 25-MAR-2003  
CC to correct FN field.)  
XX

SQ Sequence 395 AA;  
Query Match 100.0%; Score 44; DB 2; Length 395;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
Db 90 KPSSPPE 97

## RESULT 46

AA66262  
ID AAR66262 standard; protein; 395 AA.

AC AAR66262;

DT 25-MAR-2003 (revised)  
DT 17-AUG-1995 (first entry)

DE Bifunctional urokinase variant M29.

XX  
KM fibrinolysis; thrombin inhibition; thrombolytic; anti-thrombotic;  
KM urokinase; variant; mutein.

OS Synthetic.

XX  
FH Key Location/Qualifiers  
FT Region 1..365  
/label= M4  
/note= "unglycosylated prourokinase (Ser47-Leu411)"

FT Disulfide-bond 4..85

FT Disulfide-bond 25..67

FT Disulfide-bond 56..80

FT Disulfide-bond 102..233

FT Disulfide-bond 143..159

FT Disulfide-bond 151..222

FT Disulfide-bond 247..316

FT Disulfide-bond 279..295

FT Disulfide-bond 306..334

FT Region /label= X1

FT Region 372..395

FT Region /label= Y1

PN DE4323754-C1.

PD 01-DEC-1994.

PF 15-JUL-1993; 93DE-04323754.

PR 15-JUL-1993; 93DE-04323754.

XX (CHEF ) GRUENENTHAL GMBH.

XX Steffens GJ, Wendt S, Schneider J, Heinzel-Wieland R;  
XX Saunders DJ;  
XX WPI; 1995-015191/03.

XX New bifunctional urokinase derivs and related plasmids - with improved  
XX fibrinolytic and thrombin inhibiting activities, for treating cardiac and  
XX cerebral infarct, pulmonary embolism, etc.  
XX Example 1; Page 11 and Fig 1; 34pp; German.  
XX

XX Bifunctional urokinase derivatives corresponding to the formula M4-X1-Y1  
CC are claimed (see features table). Sequences AAR66244-R66266 are specific  
CC examples of such derivs. which have both improved fibrinolytic and  
CC thrombin-inhibiting activities, compared to known plasminogen activators  
CC or thrombin inhibitors. The proteins are useful as thrombolytic agents,  
CC e.g. for treatment of arterial occlusions, deep vein thrombosis, cardiac  
CC and cerebral infarction and pulmonary embolism. (Updated on 25-MAR-2003  
CC to correct PN field.)  
XX  
SQ Sequence 395 AA;  
XX  
Query Match 100.0%; Score 44; DB 2; Length 395;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KPSSPPEE 8  
DB 90 KPSSPPEE 97  
XX  
RESULT 47  
AAR66246  
ID AAR66246 standard; protein; 396 AA.  
XX  
AC AAR66246;  
XX  
DT 25-MAR-2003 (revised)  
DT 17-AUG-1995 (first entry)  
XX  
DE Bifunctional urokinase variant M13.  
XX  
KW fibrinolysis; thrombin inhibition; thrombolytic; anti-thrombotic;  
KW urokinase; variant; mutain.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Region 1..365  
FT /label= M4  
FT /note= "unglycosylated prourokinase (Ser47-Leu411)"  
FT Disulfide-bond 4..85  
FT Disulfide-bond 25..67  
FT Disulfide-bond 56..80  
FT Disulfide-bond 102..233  
FT Disulfide-bond 143..159  
FT Disulfide-bond 151..222  
FT Disulfide-bond 247..316  
FT Disulfide-bond 279..295  
FT Disulfide-bond 306..334  
FT Disulfide-bond 356..377  
FT /label= X1  
FT /label= X1  
FT Region 378..396  
FT /label= Y1  
XX  
PD DE4323754-C1.  
XX  
PD 01-DEC-1994.  
XX  
PF 15-JUL-1993; 93DE-04323754.  
XX  
PR 15-JUL-1993; 93DE-04323754.  
XX  
PA (CHRP) GRUENTHAL GMBH.  
XX  
PI Steffens GJ, Wendt S, Schneider J, Heinzel-Mieland R;  
PI Sanders DJ;  
XX  
DR WPI; 1995-015191/03.  
XX  
PT New bifunctional urokinase derivs and related plasmids - with improved  
PT fibrinolytic and thrombin inhibiting activities, for treating cardiac and  
PT cerebral infarct, pulmonary embolism, etc.

XX Example 1; Page 10 and Fig 1; 34pp; German.  
PS  
XX Bifunctional urokinase derivatives corresponding to the formula M4-X1-Y1  
CC are claimed (see features table). Sequences AAR66244-R66266 are specific  
CC examples of such derivs. which have both improved fibrinolytic and  
CC thrombin-inhibiting activities, compared to known plasminogen activators  
CC or thrombin inhibitors. The proteins are useful as thrombolytic agents,  
CC e.g. for treatment of arterial occlusions, deep vein thrombosis, cardiac  
CC and cerebral infarction and pulmonary embolism. (Updated on 25-MAR-2003  
CC to correct PN field.)  
XX  
SQ Sequence 396 AA;  
XX  
Query Match 100.0%; Score 44; DB 2; Length 396;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KPSSPPEE 8  
DB 90 KPSSPPEE 97  
XX  
RESULT 48  
AAR1828  
ID AAR1828 standard; protein; 397 AA.  
XX  
AC AAR1828;  
XX  
DT 08-JUL-1991 (first entry)  
XX  
DE FB-FB-UK fusion conjugate.  
XX  
KW Fibrin-binding protein; fibrinolysis; intravascular thrombi; fibrinogen;  
KW urokinase; fusion protein.  
XX  
OS Staphylococcus aureus.  
XX  
FH Key Location/Qualifiers  
FT Peptide 2..61  
FT /label= FB monomer  
FT Peptide 62..119  
FT /label= FB monomer  
FT Peptide 120  
FT /label= linker  
FT Peptide 121..397  
FT /label= human urokinase-B chain  
XX  
PN US5011686-A.  
XX  
PD 30-APR-1991.  
XX  
PF 15-NOV-1989; 89US-00437769.  
XX  
PR 21-SEP-1987; 87US-00099242.  
XX  
PA (CREA-) CREATIVE BIOMOLEC.  
XX  
PI Pang RHL;  
XX  
DR WPI; 1991-140198/19.  
DR N-PSDB; AAQ11650.  
XX  
PT Imparting injectable fibrinolytic agent - with affinity for intravascular  
PT thrombus, by linking agent to fibrin binding domain.  
XX  
PS Disclosure; Fig 3; 18pp; English.  
XX  
CC The conjugate comprises an FB-FB dimer linked via a Ser residue to the B  
CC chain of human urokinase. The FB fragment has selective affinity for  
CC fibrin, low affinity for fibrinogen, and minimal immunogenicity,  
CC imparting thrombus-targeting capability. See also AAR11821 and AAR11829  
XX

SQ Sequence 397 AA;  
 Query Match 100.0%; Score 44; DB 2; Length 397;  
 Best Local Similarity 100.0%; Pred. No. 99;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
 |||||  
 121 KPSSPPE 128

Db

RESULT 49  
 AAR6248  
 ID AAR6248 standard; protein; 397 AA.  
 AC AAR6248;  
 XX  
 XX 25-MAR-2003 (revised)  
 DT 17-AUG-1995 (first entry)  
 XX  
 XX Bifunctional urokinase variant M15.  
 DE  
 XX fibrinolysis; thrombin inhibition; thrombolytic; anti-thrombotic;  
 KM urokinase; variant; mutein.  
 XX  
 OS Synthetic.  
 XX  
 FH Key  
 FT Region  
 FT 1..365  
 FT /label= M4  
 FT /note="unglycosylated prourokinase(Ser47-Leu411)"  
 FT Disulfide-bond 4..85  
 FT Disulfide-bond 25..67  
 FT Disulfide-bond 56..80  
 FT Disulfide-bond 102..233  
 FT Disulfide-bond 143..159  
 FT Disulfide-bond 151..222  
 FT Disulfide-bond 247..316  
 FT Disulfide-bond 279..295  
 FT Disulfide-bond 306..334  
 FT Disulfide-bond 366..378  
 FT Region /label= X1  
 FT /label= X1  
 FT /label= X1  
 FT Region /label= Y1

DE4323754-CL.  
 PD 01-DEC-1994.  
 XX  
 XX 15-JUL-1993; 93DE-04323754.  
 PF  
 XX  
 XX 15-JUL-1993; 93DE-04323754.  
 PR  
 XX  
 XX (CHEF) GRUENENTHAL GMBH.  
 PA  
 XX Steffens GJ, Wendt S, Schneider J, Heinzel-Wieland R;  
 PI Saunders DJ;  
 XX  
 XX WPI; 1995-015191/03.  
 DR  
 XX  
 XX New bifunctional urokinase derivs and related plasmids - with improved  
 PT fibrinolytic and thrombin inhibiting activities, for treating cardiac and  
 PT cerebral infarct, pulmonary embolism, etc.  
 XX  
 XX Example 1; Page 10 and Fig 1; 34pp; German.  
 PS  
 XX Bifunctional urokinase derivatives corresponding to the formula M4-X1-Y1  
 CC are claimed (see features table). Sequences AAR6244-R66266 are specific  
 CC examples of such derivs. which have both improved fibrinolytic and  
 CC thrombin-inhibiting activities, compared to known plasminogen activators  
 CC or thrombin inhibitors. The proteins are useful as thrombolytic agents,  
 CC e.g. for treatment of arterial occlusions, deep vein thrombosis, cardiac  
 CC and cerebral infarction and pulmonary embolism. (Updated on 25-MAR-2003

CC to correct FN field.)  
 XX  
 SQ Sequence 397 AA;  
 Query Match 100.0%; Score 44; DB 2; Length 397;  
 Best Local Similarity 100.0%; Pred. No. 99;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
 |||||  
 90 KPSSPPE 97

Db

RESULT 50  
 AAM13637  
 ID AAM13637 standard; protein; 401 AA.  
 AC AAM13637;  
 XX  
 XX 04-JUN-1997 (first entry)  
 DT  
 XX  
 XX Human prourokinase variant lacking EGF domain loop 3.  
 DE  
 XX  
 XX Human prourokinase; hPUK; variant; half-life; increase; EGF;  
 KM epidermal growth factor domain; deletion; thrombolytic; fibrinolysis.  
 XX  
 XX Homo sapiens.  
 OS Synthetic.  
 XX  
 FH Key  
 FT Region  
 FT 1..32  
 FT /note="residues 1-32 of native hPUK"  
 FT Region 33..401  
 FT /note="residues 43-411 of native hPUK"  
 FT Misc-difference 167  
 FT /note="corresponds to TAC codon"

EP398361-A.  
 PD 22-NOV-1990.  
 XX  
 XX 18-MAY-1990; 90EP-00109472.  
 PF  
 XX  
 XX 18-MAY-1989; 89UP-00126433.  
 PR 22-FEB-1990; 90JP-00042020.  
 XX  
 XX (GREC) GREEN CROSS CORP.  
 PA  
 XX Tanabe T, Amatsuji Y, Kasai S, Hirose M, Morita M, Kawabe H;  
 PI Aitamura H;  
 XX  
 XX WPI; 1990-350146/47.  
 DR N-PSDB; AAT61674.  
 XX  
 XX Human pro-urokinase variants - deficient in loop regions of epidermal  
 PT growth factor, showing long blood half-life, as fibrinolytic agent.  
 PT  
 XX  
 XX Claim 11; Page; 22pp; English.  
 PS  
 XX New variants of human prourokinase (hPUK) comprise a hPUK deficient in  
 CC (1) at least part of the first loop region of the epidermal growth factor  
 CC (EGF) domain; (11) at least part of the first loop and at least part of  
 CC the second loop; or (111) at least part of the third loop. The hPUK  
 CC variants show an increased blood half-life comparable to that of the  
 CC whole EGF domain-deficient hPUK variant and urokinase while retaining the  
 CC same properties as those of hPUK. They have potent thrombolytic activity  
 CC and very little tendency to cause spontaneous bleeding. The present  
 CC sequence represents a specific variant of hPUK which lacks EGF domain  
 CC loop 3; the sequence does not appear in the specification and has been  
 CC created using the wild-type hPUK sequence and the junction sequence after  
 CC deletion, both of which are given (in Fig 1 and on page 8, respectively)  
 XX  
 SQ Sequence 401 AA;



Query Match 100.0%; Score 44; DB 2; Length 401;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8  
|||  
Db 126 KPSSPPEE 133

Search completed: June 18, 2004, 12:58:45  
Job time : 71 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: June 18, 2004, 12:52:33 ; Search time 11 Seconds  
(without alignments)  
37.869 Million cell updates/sec

Title: US-10-655-201-2

Perfect score: 44

Sequence: 1 KPSPPEB 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

No.	Score	Query Match	Length	ID	Description
1	44	100.0	431	1 UROK_HUMAN	P00749 homo sapien
2	44	100.0	433	1 UROK_PAPCY	P16227 papio cynoc
3	36	81.8	418	1 PEDF_HUMAN	P36955 homo sapien
4	36	81.8	892	1 YB96_HUMAN	O96km6 homo sapien
5	36	81.8	1088	1 PIGO_HUMAN	O8te88 homo sapien
6	36	81.8	1358	1 SIR4_YEAST	P11978 saccharomyc
7	36	81.8	1806	1 CALB_HUMAN	P12107 homo sapien
8	35	79.5	249	1 IAP2_NPVAC	P41454 autographa
9	35	79.5	522	1 PINC_CANFA	O28275 canis fam1
10	35	79.5	522	1 PINC_HORSE	O28377 equus cabal
11	35	79.5	754	1 ASRH_BOVIN	O28056 bos taurus
12	35	79.5	784	1 K6PP_HUMAN	O01813 mus sapien
13	35	79.5	784	1 K6PP_MOUSE	O9wua3 mus musculu
14	35	79.5	786	1 K6PP_RAT	P47860 rattus norv
15	35	79.5	791	1 K6PP_RABIT	P47859 oryctolagus
16	35	79.5	1240	1 DPOL_HSV21	P07918 herpes simp
17	35	79.5	1729	1 TABP_HUMAN	O9c0c2 homo sapien
18	35	79.5	1878	1 BAZA_HUMAN	O9u1f9 homo sapien
19	35	79.5	1912	1 PTPD_HUMAN	P23468 homo sapien
20	35	79.5	1948	1 PTNS_HUMAN	O13332 homo sapien
21	35	79.5	2265	1 PINC_BOVIN	P07589 bos taurus
22	35	79.5	2386	1 PINC_HUMAN	P02751 homo sapien
23	35	79.5	2477	1 PINC_MOUSE	P11276 mus musculu
24	35	79.5	2477	1 PINC_RAT	P04937 rattus norv
25	34	77.3	382	1 GBP3_ARATH	P42776 arabidopsi
26	34	77.3	462	1 MM12_MOUSE	P34960 mus musculu
27	34	77.3	499	1 GAG_HVLA2	P05887 human immun
28	34	77.3	499	1 GAG_HVLA2	P04591 human immun
29	34	77.3	499	1 GAG_HVLA3	P12464 human immun
30	34	77.3	499	1 GAG_HVLA4	O70622 human immun
31	34	77.3	499	1 GAG_HVLA5	P12463 human immun
32	34	77.3	500	1 GAG_HVLA6	P05890 human immun
33	34	77.3	501	1 GAG_HVLA2	P03349 human immun

34	34	77.3	503	1 GAG_HVLA7	P20873 human immun
35	34	77.3	506	1 GAG_HVLA8	P05888 human immun
36	34	77.3	511	1 GAG_HVLA9	P03347 human immun
37	34	77.3	511	1 GAG_HVLA5	P04593 human immun
38	34	77.3	511	1 GAG_HVLA6	P03348 human immun
39	34	77.3	511	1 GAG_HVLA7	P03350 human immun
40	34	77.3	553	1 MIS_MOUSE	P49000 rattus norv
41	34	77.3	555	1 MIS_MOUSE	P27106 mus musculu
42	34	77.3	598	1 NR42_HUMAN	P43354 homo sapien
43	34	77.3	598	1 NR42_MOUSE	O06219 mus musculu
44	34	77.3	598	1 NR42_RAT	O07917 rattus norv
45	34	77.3	615	1 YJ13_SCHPO	O13681 schizosacch
46	34	77.3	886	1 CN4A_HUMAN	P27815 homo sapien
47	34	77.3	928	1 RSC1_YEAST	P53236 saccharomyc
48	34	77.3	959	1 G2D1_HUMAN	O9u1f9 h general t
49	34	77.3	1080	1 MTRB_MOUSE	P59759 mus musculu
50	34	77.3	1088	1 MTRB_HUMAN	O9u1f7 mus sapien
51	34	77.3	1104	1 G2D1_MOUSE	O9j157 mus sapien
52	34	77.3	1567	1 RMI_DROME	O9v7h4 drosophila
53	33	75.0	153	1 GSPG_BRWCH	P31585 erwina chr
54	33	75.0	180	1 DADR_RABIT	O02664 oryctolagus
55	33	75.0	245	1 PCYA_ANASP	O93c10 anabaena sp
56	33	75.0	366	1 CD44_BOVIN	O29423 bos taurus
57	33	75.0	375	1 KORA_METTH	O27112 methanobact
58	33	75.0	393	1 STAD_SOLCO	O41319 solanum com
59	33	75.0	397	1 DJA4_HUMAN	O8wv22 homo sapien
60	33	75.0	397	1 DJA4_MOUSE	O9jmc3 mus musculu
61	33	75.0	399	1 STAD_SPIOL	P28645 spiracle ol
62	33	75.0	411	1 IF2G_PPRAB	O9v1g0 pyrococcus
63	33	75.0	411	1 IF2G_PPRFU	O8u082 pyrococcus
64	33	75.0	411	1 IF2G_PPRFU	O59410 pyrococcus
65	33	75.0	424	1 IF2G_PPRFU	O29663 archaeoglob
66	33	75.0	506	1 GAG_SIVM1	P05894 simian immu
67	33	75.0	514	1 CSA_DIEMI	P08796 dicystosteli
68	33	75.0	640	1 TERM_ADE07	P03270 human adeno
69	33	75.0	830	1 MCM2_SCHPO	P40377 schizosacch
70	33	75.0	1093	1 PIGO_MOUSE	O9j316 mus musculu
71	33	75.0	1097	1 KFLD_RAT	O35787 rattus norv
72	33	75.0	1296	1 ASAI_ENTPA	P17953 enterococcu
73	33	75.0	3644	1 MINT_MOUSE	O62504 mus musculu
74	33	75.0	5262	1 ML12_HUMAN	O14686 homo sapien
75	32	72.7	138	1 LACR_HUMAN	O9g248 homo sapien
76	32	72.7	145	1 SELM_HUMAN	O8wv29 homo sapien
77	32	72.7	145	1 SELM_MOUSE	O8vnc3 mus musculu
78	32	72.7	174	1 PTRY_SPIOL	P38365 spiracle ol
79	32	72.7	243	1 FCBB_RAT	P13386 rattus norv
80	32	72.7	243	1 NUKS_HUMAN	O9h1e3 homo sapien
81	32	72.7	247	1 SURE_CHRVO	O7nxy1 chromobacte
82	32	72.7	252	1 TRT3_COTJA	P06398 coturnix co
83	32	72.7	262	1 TRT3_CHICK	P12620 gallus gall
84	32	72.7	283	1 EXTN_SORBI	P24152 sorghum bic
85	32	72.7	377	1 SX18_MOUSE	P43680 mus musculu
86	32	72.7	430	1 SYS_THETH	P34945 thermus the
87	32	72.7	431	1 TOLB_YERPE	O8z921 yerinia pe
88	32	72.7	446	1 PIV2_ADE40	P48752 human adeno
89	32	72.7	452	1 ETV6_HUMAN	P41312 homo sapien
90	32	72.7	475	1 RBL_PULAU	O43036 plumbago au
91	32	72.7	476	1 KE4_MOUSE	O31125 mus musculu
92	32	72.7	538	1 CN93_HUMAN	O9h972 homo sapien
93	32	72.7	539	1 Y793_TREPA	O83791 treponema p
94	32	72.7	549	1 TEGU_HCMV	P07387 human cytom
95	32	72.7	562	1 SVK_AERPE	O9yft9 aeropyrum p
96	32	72.7	632	1 FAN2_HUMAN	O9y566 homo sapien
97	32	72.7	640	1 HKR1_HUMAN	P10072 homo sapien
98	32	72.7	652	1 CD93_HUMAN	O9npv3 homo sapien
99	32	72.7	685	1 SNK_HUMAN	O9nyy3 homo sapien
100	32	72.7	783	1 YAYB_SCHPO	O10218 schizosacch
101	32	72.7	875	1 UBP7_SCHPO	O9p7b5 schizosacch
102	32	72.7	941	1 DNAB_PROMR	O30477 rhodothermu
103	32	72.7	953	1 LDM1_ARATH	O38796 arabidopsi
104	32	72.7	1130	1 SN3B_HUMAN	O75182 homo sapien
105	32	72.7	1234	1 JMW_MOUSE	O62315 mus musculu
106	32	72.7	1266	1 JMW_HUMAN	O92833 homo sapien

107	32	72.7	1281	1	IP35_MOUSE	Q9329	mus musculus	180	31	70.5	805	1	TACT_HUMAN	Q75410	homo sapien
108	32	72.7	1426	1	AI0D_HUMAN	Q9241	homo sapien	181	31	70.5	820	1	YNH4_CAEEL	P32742	caenorhabd
109	32	72.7	2701	1	IP35_RAT	P2995	rattus norv	182	31	70.5	860	1	AREA_PENRO	O13508	penicillium
110	32	72.7	3130	1	PD02_HUMAN	O6067	homo sapien	183	31	70.5	862	1	Z409_HUMAN	Q9296	homo sapien
111	31	70.5	78	1	YDZH_BACSU	O31496	baacillus su	184	31	70.5	865	1	NRPA_PENUR	Q9266	penicillium
112	31	70.5	82	1	HPIS_MARPU	P59860	marichromac	185	31	70.5	872	1	SCDS_YEAST	P34758	saccharomyc
113	31	70.5	83	1	HPIS_CHRGR	P00262	chromochrom	186	31	70.5	911	1	CAPA_MOUSE	Q9QW40	mus musculus
114	31	70.5	83	1	HPIS_THETI	P80176	thermochrom	187	31	70.5	928	1	PTP3_YEAST	P40048	saccharomyc
115	31	70.5	87	1	NXL4_BUNMU	O12961	bungarus mu	188	31	70.5	932	1	VGLE_HSVBP	P17471	bovine hepr
116	31	70.5	94	1	KV11_RABIT	P01662	oryctolagus	189	31	70.5	932	1	VGLE_HSVBP	P12640	bovine hepr
117	31	70.5	101	1	VF17_VAVU	P33875	variola vir	190	31	70.5	956	1	NU11_MAGGR	O01168	magnaporthe
118	31	70.5	122	1	HPIS_CHRVT	P00260	chromatium	191	31	70.5	971	1	AREA_GIBFU	P78668	gibberella
119	31	70.5	127	1	LEC_GALNI	P30617	galactinus n	192	31	70.5	1034	1	ENTK_PIG	P98074	sus scrofa
120	31	70.5	155	1	GCSH_ARATH	P25855	arabidopsis	193	31	70.5	1036	1	NIT2_NEUCR	P19212	neurospora
121	31	70.5	166	1	GCSI_ARATH	Q91410	arabidopsis	194	31	70.5	1043	1	Y483_CHLPU	Q92688	chlamydia p
122	31	70.5	184	1	PEPI_HUMAN	Q8N49	homo sapien	195	31	70.5	1132	1	YKKS_YEAST	P34250	saccharomyc
123	31	70.5	228	1	VSPA_LACMU	P33569	lachesia mu	196	31	70.5	1141	1	MYPS_HUMAN	Q00872	homo sapien
124	31	70.5	227	1	YGHS_ECOLI	Q46843	escherichia	197	31	70.5	1210	1	AT19_MOUSE	P59509	mus musculus
125	31	70.5	257	1	VSPC_TRIGA	O13062	trimeresuru	198	31	70.5	1389	1	LTHS_MOUSE	Q8CG18	mus musculus
126	31	70.5	252	1	BGSX_ACEXY	Q9W659	acetobacter	199	31	70.5	1427	1	ZFH2_HUMAN	Q9CG41	homo sapien
127	31	70.5	297	1	PEXC_DROME	Q9VPS5	acrobacter	200	31	70.5	1440	1	SYEP_HUMAN	P07814	homo sapien
128	31	70.5	304	1	DEPC_BACPF	P94312	baacillus ps	201	31	70.5	1713	1	LTHL_MOUSE	Q8CG19	mus musculus
129	31	70.5	307	1	COPE_BOVIN	O28104	bos taurus	202	31	70.5	1820	1	STP9_HUMAN	Q92D5	mus sapien
130	31	70.5	307	1	COPE_CRIGR	Q60445	cricetulus	203	31	70.5	1942	1	HEI2_HUMAN	P42674	homo sapien
131	31	70.5	307	1	COPE_HUMAN	O14579	homo sapien	204	31	70.5	2065	1	T2D1_DROME	P51123	drosophila
132	31	70.5	311	1	ATH6_ARATH	P46668	arabidopsis	205	31	70.5	2763	1	TEGU_VZVD	Q92728	varicella-z
133	31	70.5	314	1	YDQC_SCHPO	O14204	schizosach	206	31	70.5	2774	1	MAPA_RAT	P34926	rattus norv
134	31	70.5	316	1	APB_BOVIN	Q03247	bos taurus	207	31	70.5	2805	1	MLA3_MOUSE	P78559	mus sapien
135	31	70.5	328	1	VMSA_HPBUD	P3145	duck hepati	208	31	70.5	4903	1	PCLO_MOUSE	Q8BH4	mus musculus
136	31	70.5	329	1	GC3_MOUSE	P22436	mus musculus	209	31	70.5	5038	1	PCLO_MOUSE	Q9QX7	mus musculus
137	31	70.5	373	1	MEB1_HUMAN	Q9UB11	homo sapien	210	31	70.5	5085	1	PCLO_RAT	Q9JX86	rattus norv
138	31	70.5	373	1	STAD_GOSHI	Q42770	gossypium h	211	31	70.5	5120	1	PCLO_CHICK	Q9P36	gallus gall
139	31	70.5	388	1	GC3M_MOUSE	P33987	mus musculus	212	31	70.5	5147	1	PCLO_HUMAN	Q9Y670	homo sapien
140	31	70.5	398	1	STAD_BRANA	P29108	brassica na	213	31	68.2	48	1	RBL_PINS	P81080	pinus pinas
141	31	70.5	399	1	SP2_HUMAN	O81W5	homo sapien	214	31	68.2	53	1	RBL_MALDO	P31194	malus domes
142	31	70.5	405	1	GDPB_MOUSE	Q92144	mus musculus	215	31	68.2	54	1	RBL_GERBO	P31188	geum borisi
143	31	70.5	415	1	EF1G_YEAST	P29547	saccharomyc	216	31	68.2	54	1	RBL_ICAMA	P31190	icactina man
144	31	70.5	416	1	YHR5_YEAST	P38823	saccharomyc	217	31	68.2	54	1	RBL_ILEAQ	P31191	illex aquilo
145	31	70.5	463	1	D2DR_FUGRU	P53433	fugu rubrip	218	31	68.2	54	1	RBL_ILECI	P31192	illex aquilo
146	31	70.5	465	1	SNX8_HUMAN	Q9Y5X2	homo sapien	219	31	68.2	54	1	RBL_ILEPE	Q07049	illex pernyi
147	31	70.5	480	1	DDC_CAVPO	P22781	cavia porce	220	31	68.2	54	1	RBL_MAGLI	P31193	magnolia il
148	31	70.5	480	1	IRXI_MOUSE	P81068	mus musculus	221	31	68.2	54	1	RBL_MERAN	P31195	mercurialis
149	31	70.5	480	1	KLPA_RAT	O08876	rattus norv	222	31	68.2	54	1	RBL_RHACA	P31196	rhizanthus cat
150	31	70.5	482	1	PO24_POPJA	O03275	popillia ja	223	31	68.2	54	1	RBL_RIBAM	P31199	ribes ameri
151	31	70.5	486	1	MEC2_HUMAN	P51608	homo sapien	224	31	68.2	57	1	RBL_BOXSE	P31180	buxus sempe
152	31	70.5	496	1	GTR3_CHICK	P28568	gallus gall	225	31	68.2	57	1	RBL_CAMSI	P31181	camellia si
153	31	70.5	505	1	FXO4_HUMAN	P98177	homo sapien	226	31	68.2	58	1	RBL_EUOBU	P31185	eunymus bu
154	31	70.5	520	1	RXR8_MOUSE	P28717	homo sapien	227	31	68.2	58	1	RBL_EUOMA	P31186	eunymus bu
155	31	70.5	527	1	GI9P_HUMAN	P14314	homo sapien	228	31	68.2	58	1	RBL_EUPCH	P31187	euphorbia c
156	31	70.5	534	1	APG_ARATH	P40602	arabidopsis	229	31	68.2	58	1	RBL_WEISY	P31200	weilmannia
157	31	70.5	572	1	CAF8_MOUSE	Q94057	mus musculus	230	31	68.2	58	1	RBL_ROSDA	P31196	nemopanthus
158	31	70.5	575	1	TERM_ADEGI	O64752	avian adeno	231	31	68.2	143	1	RBL_NEMMU	P50550	homo sapien
159	31	70.5	601	1	YFK5_SCHPO	P87132	schizosach	232	31	68.2	158	1	UBCT_HUMAN	O09181	mesocricetu
160	31	70.5	611	1	IF4B_HUMAN	P23588	homo sapien	233	31	68.2	158	1	UBCT_MESAU	P02810	homo sapien
161	31	70.5	656	1	DCHS_RAT	P14523	rattus norv	234	31	68.2	166	1	PRPC_HUMAN	Q08195	homo sapien
162	31	70.5	658	1	Y285_HUMAN	O9144	sus scrofa	235	31	68.2	173	1	FALP_HUMAN	Q8CT5	homo sapien
163	31	70.5	668	1	SG1_PIG	O95447	homo sapien	236	31	68.2	182	1	CXZ1_MOUSE	Q9CG40	mus musculus
164	31	70.5	670	1	CUL3_HUMAN	P50478	gallus gall	237	31	68.2	211	1	RCN1_YEAST	P36054	saccharomyc
165	31	70.5	682	1	AMPH_CHICK	O08838	rattus norv	238	31	68.2	234	1	PRPW_HUMAN	P10161	homo sapien
166	31	70.5	683	1	AMPH_RAT	P10090	drosophila	239	31	68.2	243	1	YK42_MYCTU	O53392	mycobacteri
167	31	70.5	687	1	WHIT_DROME	O9U16	homo sapien	240	31	68.2	247	1	PRB4_HUMAN	P10163	homo sapien
168	31	70.5	695	1	AMPH_HUMAN	P49418	homo sapien	241	31	68.2	249	1	APXI_ARATH	O05431	arabidopsis
169	31	70.5	744	1	KCN3_PIG	P58332	sus scrofa	242	31	68.2	249	1	APXI_PEA	P48534	pisum sativ
170	31	70.5	731	1	KCN3_MOUSE	P70605	rattus norv	243	31	68.2	251	1	PRP2_HUMAN	O08873	caenorhabd
171	31	70.5	732	1	KCN3_RAT	Q9UP6	homo sapien	244	31	68.2	273	1	HMD1_CAEEL	P10162	homo sapien
172	31	70.5	733	1	SAB2_HUMAN	O8V14	mus musculus	245	31	68.2	276	1	PRPL_HUMAN	Q00165	homo sapien
173	31	70.5	733	1	SAB2_MOUSE	O9UG16	homo sapien	246	31	68.2	279	1	HAK1_HUMAN	O9B011	cericobebus
174	31	70.5	736	1	KCN3_HUMAN	O12797	homo sapien	247	31	68.2	280	1	TNFE_CERTO	O9WY16	macaca mula
175	31	70.5	757	1	ASPH_HUMAN	P12757	homo sapien	248	31	68.2	281	1	TNFE_HACMU	P48023	homo sapien
176	31	70.5	763	1	IF2C_PORPU	P1257	porphyra pu	249	31	68.2	285	1	TNVO_CANPA	P18174	cantis fami
177	31	70.5	773	1	YOD3_CAEEL	P4595	caenorhabd	250	31	68.2	285	1	FX16_HUMAN	O8IX29	homo sapien
178	31	70.5	786	1	CG2F_HUMAN	P41002	homo sapien	251	31	68.2	292	1	VGIG_HRSV2	P27021	human tespi
179	31	70.5	802	1	NAB3_YEAST	P38996	saccharomyc	252	31	68.2	297	1	VGIG_HRSV2		

253	30	68.2	297	1	VGLG_HRSV3	P27022 human respi	326	30	68.2	443	1	RBL_VERBO	P36490 verbera bon
254	30	68.2	298	1	VGLG_HRSVA	P03423 human respi	327	30	68.2	443	1	RBL_VITCA	Q05994 villareia c
255	30	68.2	298	1	VGLG_HRSVA	P20895 human respi	328	30	68.2	443	1	SOX3_HUMAN	P41225 homo sapien
256	30	68.2	309	1	SOX3_XENLA	P55863 xenopus lae	329	30	68.2	444	1	RBL_GINBT	P48704 ginkgo bilo
257	30	68.2	315	1	RPO4_THETH	Q929h6 thetmus the	330	30	68.2	444	1	RBL_WELMT	P48719 welwitschia
258	30	68.2	318	1	MYOD_RAT	Q02346 rattus norv	331	30	68.2	445	1	GD1B_HUMAN	P50395 homo sapien
259	30	68.2	323	1	GC_RABIT	P01870 corytolagus	332	30	68.2	445	1	RBL_CALDI	P36483 callicarpa
260	30	68.2	327	1	Z444_HUMAN	Q8n0y2 homo sapien	333	30	68.2	446	1	RBL_EXAAF	Q05989 exacum affi
261	30	68.2	331	1	PRP1_HUMAN	P04280 homo sapien	334	30	68.2	447	1	RBL_CAMLE	Q05694 camassia le
262	30	68.2	334	1	FX16_MOUSE	Q9qzms mus musculu	335	30	68.2	447	1	RBL_CONTR	Q05988 convolvulus
263	30	68.2	346	1	PF16_ECOLI	P06128 escherichia	336	30	68.2	447	1	RBL_LIGUV	Q05991 ligustrum v
264	30	68.2	349	1	US30_HCMVA	P09706 human cytom	337	30	68.2	448	1	RBL_EUCTU	P28414 eucryphia l
265	30	68.2	375	1	SOX3_MOUSE	P51874 mus musculu	338	30	68.2	449	1	RBL_HIPRI	P31189 hippocratea
266	30	68.2	376	1	PEXE_MOUSE	Q939a0 mus musculu	339	30	68.2	449	1	RBL_SALPL	P31202 salacia pal
267	30	68.2	390	1	STAD_OLEEU	Q43593 olea europae	340	30	68.2	449	1	RBL_ZAMZA	Q06681 zamoculcaes
268	30	68.2	390	1	STAD_OLEUA	Q40731 oleya sativ	341	30	68.2	450	1	FOKE_CHLUP	Q92768 c foliate sy
269	30	68.2	394	1	RBL_ALIPL	P34767 allama plan	342	30	68.2	450	1	RBL_CRAMA	P28395 cedrusula ma
270	30	68.2	394	1	RBL_BARIO	Q05795 barclaya lo	343	30	68.2	450	1	RBL_SBDPU	P28455 sedum rubro
271	30	68.2	394	1	RBL_BRASC	Q05796 braesenia sc	344	30	68.2	453	1	RBL_ASPIA	Q31809 asperula la
272	30	68.2	394	1	RBL_CABCA	Q05797 cabomba car	345	30	68.2	453	1	RBL_CRYAN	Q31946 crucianella
273	30	68.2	394	1	RBL_CERDE	Q05798 ceratophyll	346	30	68.2	453	1	RBL_CRUGL	Q31992 cruciata gl
274	30	68.2	394	1	RBL_EURPE	Q05799 euryale fer	347	30	68.2	453	1	RBL_GALAL	Q32255 galium albu
275	30	68.2	394	1	RBL_NELLU	Q05800 nelumbo lut	348	30	68.2	453	1	RBL_GALAP	Q32256 galium apar
276	30	68.2	394	1	RBL_NUPVA	Q05801 nuphar vari	349	30	68.2	453	1	RBL_GALCO	Q32271 galium cors
277	30	68.2	394	1	RBL_NYMOD	Q05802 nympheae od	350	30	68.2	453	1	RBL_GALEL	Q32283 galium elon
278	30	68.2	394	1	RBL_VICCC	Q05803 victoria cr	351	30	68.2	453	1	RBL_GALLU	Q32303 galium lucu
279	30	68.2	396	1	STAD_HILAN	Q06456 helianthus	352	30	68.2	453	1	RBL_GALPA	Q32345 galium patu
280	30	68.2	398	1	C1W4_MOUSE	Q08454 mus musculu	353	30	68.2	453	1	RBL_GALPR	Q32344 galium patl
281	30	68.2	399	1	STAS_BRANA	Q01771 braesica na	354	30	68.2	453	1	RBL_GALSC	Q32360 galium scab
282	30	68.2	403	1	YD50_PROMM	P59919 prochloroco	355	30	68.2	453	1	RBL_HYDFO	Q32397 hydophyllum
283	30	68.2	404	1	KV83_RAT	Q63494 rattus norv	356	30	68.2	453	1	RBL_PHOST	Q32808 phopsis sty
284	30	68.2	406	1	RBL_GREBI	P43329 crepidomane	357	30	68.2	453	1	RBL_RUBTI	Q33050 rubia tinct
285	30	68.2	410	1	RBL_GLEJA	P48705 gleichenia	358	30	68.2	453	1	RBL_SHEAR	Q33062 sherardia a
286	30	68.2	411	1	STAD_SOYBN	Q42807 glycine max	359	30	68.2	453	1	RBL_VALMU	Q33574 valantia ma
287	30	68.2	411	1	VG66_HSV11	Q00154 ictaluriid h	360	30	68.2	455	1	RBL_ACAFA	P93398 acacia fatm
288	30	68.2	414	1	SX17_HUMAN	Q9h612 homo sapien	361	30	68.2	455	1	RBL_BROCO	P93388 bromnea coc
289	30	68.2	416	1	RBL_CHEBI	P43227 cheilopleur	362	30	68.2	455	1	RBL_CASDI	P93680 caesia didy
290	30	68.2	416	1	RBL_SPTMR	P36488 spigelia ma	363	30	68.2	455	1	RBL_BRVCG	Q33438 erythrina c
291	30	68.2	419	1	SX17_MOUSE	Q61473 mus musculu	364	30	68.2	455	1	RBL_LUPRA	Q92398 lupinus acil
292	30	68.2	420	1	RBL_ANEME	Q31674 anemula mexi	365	30	68.2	455	1	RBL_LUPAB	P52781 lupinus alb
293	30	68.2	426	1	RBL_HORVU	P05698 hordeum vul	366	30	68.2	455	1	RBL_LUPAL	P92396 lupinus alb
294	30	68.2	427	1	TR16_HUMAN	Q90813 homo sapien	367	30	68.2	455	1	RBL_LUPAL	P92397 lupinus alb
295	30	68.2	430	1	SYTB_MOUSE	Q90813 mus musculu	368	30	68.2	455	1	RBL_LUPAN	P52773 lupinus ang
296	30	68.2	430	1	SYTB_RAT	Q08835 rattus norv	369	30	68.2	455	1	RBL_LUPCO	P92400 lupinus cos
297	30	68.2	435	1	RBL_DROBR	P28404 drosera bur	370	30	68.2	455	1	RBL_LUPDE	P92401 lupinus den
298	30	68.2	436	1	RBL_BYALI	P28386 byalis linl	371	30	68.2	455	1	RBL_LUPMT	P92406 lupinus mic
299	30	68.2	436	1	RBL_SANCA	P28450 sanguinaria	372	30	68.2	455	1	RBL_LOPNA	P92407 lupinus nan
300	30	68.2	440	1	RBL_BAMGL	P51594 bambusa gla	373	30	68.2	455	1	RBL_LUPPO	P52775 lupinus pol
301	30	68.2	440	1	RBL_DICAN	P48701 dicksonia a	374	30	68.2	455	1	RBL_TAMIN	P93669 tamaxindus
302	30	68.2	440	1	RBL_PTYGA	Q098360 ptychomitri	375	30	68.2	457	1	RBL_PHECO	P31197 phelline co
303	30	68.2	441	1	RBL_APOCA	Q05984 apocynum ca	376	30	68.2	459	1	RBL_APRGR	P28380 apium grave
304	30	68.2	441	1	RBL_ASCCX	P36480 asclepias ca	377	30	68.2	459	1	RBL_CEPFO	P28390 cephalotus
305	30	68.2	441	1	RBL_BEGMS	P28383 begonia met	378	30	68.2	459	1	RBL_CERGU	P28391 ceratopetal
306	30	68.2	441	1	RBL_CORSA	Q05579 coriandrum	379	30	68.2	459	1	RBL_CORLA	P31183 corynocarpu
307	30	68.2	441	1	RBL_DARCA	P28398 darlingtoni	380	30	68.2	459	1	RBL_GAREL	P28416 garrya elll
308	30	68.2	441	1	RBL_DRODC	P28405 drosera dic	381	30	68.2	459	1	RBL_GRUCH	P28418 geum chiloe
309	30	68.2	441	1	RBL_DROPT	P28410 drosera pec	382	30	68.2	459	1	RBL_HRUMT	P28423 heuchera mi
310	30	68.2	441	1	RBL_FOUCP	Q05590 fouquieria	383	30	68.2	459	1	RBL_MORAL	P28431 morus alba
311	30	68.2	441	1	RBL_GLYXC	Q62970 glycyrrhiza	384	30	68.2	459	1	RBL_NYFPR	P28261 nyra frutic
312	30	68.2	441	1	RBL_HELNU	P28422 hellamphora	385	30	68.2	459	1	RBL_NYSOG	P28435 nyssa ogech
313	30	68.2	441	1	RBL_POLRE	Q05992 polemonium	386	30	68.2	459	1	RBL_PAREFI	P28437 parassia f
314	30	68.2	441	1	RBL_SYMAL	Q05993 symphoricar	387	30	68.2	459	1	RBL_PRUDO	P28445 prunus dome
315	30	68.2	441	1	RBL_VIOSO	Q05995 viola sorar	388	30	68.2	459	1	RBL_RORGO	P28448 rudidula go
316	30	68.2	443	1	RBL_ABFIF	Q78258 abies firma	389	30	68.2	459	1	RBL_SAXIN	P28452 saxifraga l
317	30	68.2	443	1	RBL_ABFIO	Q78259 abies homol	390	30	68.2	459	1	RBL_STRLC	Q36600 streptopus
318	30	68.2	443	1	RBL_ABFMR	Q78261 abies marie	391	30	68.2	459	1	RBL_CRYJA	P48696 cryptomeria
319	30	68.2	443	1	RBL_ABISA	Q78262 abies sachu	392	30	68.2	460	1	RBL_CUNTA	Q32026 cunninghami
320	30	68.2	443	1	RBL_ABIVE	Q78260 abies veitc	393	30	68.2	460	1	RBL_HALCL	Q98890 halesia car
321	30	68.2	443	1	RBL_ANTMA	Q05554 antirrhinum	394	30	68.2	463	1	RBL_GERGR	P28417 geranium gr
322	30	68.2	443	1	RBL_BOKOF	Q05965 borago ofit	395	30	68.2	463	1	RBL_PELHO	P28439 pelargonium
323	30	68.2	443	1	RBL_BUDDA	P36482 buddleja da	396	30	68.2	465	1	RBL_ACESA	P28376 acer saccha
324	30	68.2	443	1	RBL_CAHBE	Q05986 callitriche	397	30	68.2	465	1	RBL_ACTCH	P28377 actinidia c
325	30	68.2	443	1	RBL_SESIN	P36487 sesamum ind	398	30	68.2	465	1	RBL_AITAL	Q07209 atlantus a

399	30	68.2	465	1	RBL_BAURU	Q31730	bauera rubi
400	30	68.2	465	1	RBL_BURIN	P28385	bureira ina
401	30	68.2	465	1	RBL_BYRCR	P28387	byresonia c
402	30	68.2	465	1	RBL_CASFS	Q20304	casia fist
403	30	68.2	465	1	RBL_CASLI	P28389	casarina l
404	30	68.2	465	1	RBL_CLITE	Q20346	clitoria te
405	30	68.2	465	1	RBL_CORCA	P28393	cornus cana
406	30	68.2	465	1	RBL_CORFO	Q31983	cornus flor
407	30	68.2	465	1	RBL_COROB	Q32042	cornus obl
408	30	68.2	465	1	RBL_COROF	Q32042	cornus offi
409	30	68.2	465	1	RBL_COROT	Q32041	cornus obl
410	30	68.2	465	1	RBL_CYRRA	P28396	cyrilla rac
411	30	68.2	465	1	RBL_DABSP	P28397	daphniphyll
412	30	68.2	465	1	RBL_DILIN	P28400	dillenia in
413	30	68.2	465	1	RBL_EPASP	P28402	epacis sp.
414	30	68.2	465	1	RBL_EPHTW	Q32223	ephedra twe
415	30	68.2	465	1	RBL_EUCUL	P28413	eucornia ul
416	30	68.2	465	1	RBL_FRANR	P48703	fragaria an
417	30	68.2	465	1	RBL_HAMMO	P28419	hamamelis m
418	30	68.2	465	1	RBL_HUMMA	P28424	humilia bal
419	30	68.2	465	1	RBL_IDEFO	Q9XPX5	idesia poly
420	30	68.2	465	1	RBL_ILECR	P28446	ilex crenat
421	30	68.2	465	1	RBL_MANZA	P28430	manikara z
422	30	68.2	465	1	RBL_MORRU	Q32635	morris rubra
423	30	68.2	465	1	RBL_MYRCE	P28432	myrica ceri
424	30	68.2	465	1	RBL_NEPAL	P28441	nepenthes a
425	30	68.2	465	1	RBL_PLAOC	P28441	platanus oc
426	30	68.2	465	1	RBL_PLAVR	P28442	platytheca
427	30	68.2	465	1	RBL_POLCR	P28443	polygala cr
428	30	68.2	465	1	RBL_QUITN	P28446	quisqualis
429	30	68.2	465	1	RBL_RHOHI	P28447	rhododendro
430	30	68.2	465	1	RBL_SARPL	P28451	sarracenia
431	30	68.2	465	1	RBL_TROAR	P28458	trochodendr
432	30	68.2	465	1	RBL_ULMAL	Q33245	ulmus alata
433	30	68.2	465	1	RSP4_CHLRE	Q01656	chlamydomon
434	30	68.2	466	1	RBL_ADOBO	Q93000	adentium obe
435	30	68.2	466	1	RBL_ADOBO	P28378	adocxa mosch
436	30	68.2	466	1	RBL_AESPA	Q31827	aeschulus pa
437	30	68.2	466	1	RBL_APHSI	P28379	aphelandra
438	30	68.2	466	1	RBL_ASACA	P28479	asarum cana
439	30	68.2	466	1	RBL_AVECA	Q9MTE7	averrhoa ca
440	30	68.2	466	1	RBL_BARRR	P28382	barleria pr
441	30	68.2	466	1	RBL_BEAGR	Q93001	beaumontia
442	30	68.2	466	1	RBL_BERLA	P36481	berzelia la
443	30	68.2	466	1	RBL_BETNI	P28384	betula nigr
444	30	68.2	466	1	RBL_BIXOR	Q19872	bixa orella
445	30	68.2	466	1	RBL_BULAR	Q19989	bulnesia ar
446	30	68.2	466	1	RBL_CALPL	P28388	caltha palu
447	30	68.2	466	1	RBL_CERJA	Q05367	ceratidiphyll
448	30	68.2	466	1	RBL_CORNO	Q33369	cornus kous
449	30	68.2	466	1	RBL_CORNY	P28394	cornaria my
450	30	68.2	466	1	RBL_CUCPE	P28637	cucubita p
451	30	68.2	466	1	RBL_DICMU	P28401	dioneaea mus
452	30	68.2	466	1	RBL_DRIMI	P28402	drimys wint
453	30	68.2	466	1	RBL_DROBI	P28403	drosera din
454	30	68.2	466	1	RBL_DROCA	P28405	drosera cap
455	30	68.2	466	1	RBL_DROPI	P28407	drosera fil
456	30	68.2	466	1	RBL_DROUJ	P28408	droserophyllu
457	30	68.2	466	1	RBL_DRORE	P28411	drosera reg
458	30	68.2	466	1	RBL_EREMA	Q33443	eremochamu
459	30	68.2	466	1	RBL_HARGR	P28440	harpagophyt
460	30	68.2	466	1	RBL_HEDHE	P28421	hedera heli
461	30	68.2	466	1	RBL_JUSOD	P28428	justicia od
462	30	68.2	466	1	RBL_LOBSP	P28429	lobelia sp.
463	30	68.2	466	1	RBL_MOROL	P48708	morninga ole
464	30	68.2	466	1	RBL_MANDU	Q20241	nandina dom
465	30	68.2	466	1	RBL_NELCA	P28433	nelsonia ca
466	30	68.2	466	1	RBL_OXADI	P28436	oxalis dill
467	30	68.2	466	1	RBL_PINCA	P28440	pinguicula
468	30	68.2	466	1	RBL_POLSI	Q9GHP8	poliothyrsi
469	30	68.2	466	1	RBL_PROUJ	P28444	proscodaea
470	30	68.2	466	1	RBL_RUTRA	P28449	rutylra frut
471	30	68.2	466	1	RBL_SAUCE	P36486	saururus ce

472	30	68.2	466	1	RBL_SITGA	P25837	silene gall
473	30	68.2	466	1	RBL_THERPO	P28457	theopesia p
474	30	68.2	466	1	RBL_VITAE	P28460	vitis aesti
475	30	68.2	467	1	RBL_CALUS	P25829	calamus usi
476	30	68.2	467	1	RBL_CENAT	Q9GXX6	cedrus atla
477	30	68.2	467	1	RBL_ERICA	P28413	eriodictyon
478	30	68.2	467	1	RBL_HYDVI	P28425	hydrorhiza
479	30	68.2	467	1	RBL_JASUS	P28427	jasminum su
480	30	68.2	467	1	RBL_PHORE	P28462	phoenix rec
481	30	68.2	467	1	RBL_SCUBO	P28453	scutellaria
482	30	68.2	467	1	RBL_SERRE	P25836	serenoa rep
483	30	68.2	467	1	RBL_TASIN	P28456	tasmanian l
484	30	68.2	468	1	RBL_ANTHE	Q31669	anthospermum
485	30	68.2	468	1	RBL_ANTVS	Q31659	anthocercis
486	30	68.2	468	1	RBL_CAPBA	Q31551	capitulum ba
487	30	68.2	468	1	RBL_CATSP	Q33383	catesbaea s
488	30	68.2	468	1	RBL_CORAT	Q31348	cornus alie
489	30	68.2	468	1	RBL_CONGU	Q98883	conouplia
490	30	68.2	468	1	RBL_DATST	P48698	datuna stru
491	30	68.2	468	1	RBL_NOLSP	Q32659	nolana spat
492	30	68.2	468	1	RBL_PANUS	Q98668	pandorea ja
493	30	68.2	468	1	RBL_SALDI	P36485	salvia divi
494	30	68.2	468	1	RBL_SOLGR	Q33101	solandra gr
495	30	68.2	468	1	RBL_TECST	Q98671	tecoma stan
496	30	68.2	469	1	RBL_AKABI	Q07821	akania bidw
497	30	68.2	469	1	RBL_ANTIUV	Q31672	atriplex lu
498	30	68.2	469	1	RBL_ATRPA	P19160	atriplex pa
499	30	68.2	469	1	RBL_BREMA	Q31738	brexia mada
500	30	68.2	469	1	RBL_CALCA	Q31750	calycophyll

## ALIGNMENTS

RESULT 1

ID	UROC_HUMAN	STANDARD;	PRT;	431 AA.
AC	P00749; Q15844; Q16618; Q969W6;			
AT	21-JUL-1986 (Rel. 01, Created)			
DT	20-MAR-1987 (Rel. 04, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Urokinase-type plasminogen activator precursor (EC 3.4.21.73) (uPA)			
DE	(U-plasminogen activator).			
GN	PLAU.			
OS	Homo sapiens (Human).			
OS	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN				
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=05215647; PubMed=2987867;			
RA	Riccio A., Grimaldi G., Verde P., Sebastio G., Boast S., Biasi F.;			
RT	"The human urokinase-plasminogen activator gene and its promoter.";			
RT	Nucleic Acids Res. 13:2759-2771(1985).			
RN				
RP	SEQUENCE FROM N.A.			
RA	Holmes W.E., Pennica D., Blaber M., Rey M.W., Guenzler W.A.,			
RT	Steffens G.J., Heyneker H.L.;			
RT	"Cloning and expression of the gene for pro-urokinase in Escherichia			
RT	coli.";			
RL	Biotechnology 3:923-929(1985).			
RN				
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=06056954; PubMed=2415429;			
RA	Nagai M., Hiramatsu R., Kaneda T., Hayasuke N., Arimura H.,			
RT	Nishida M., Suyama T.;			
RT	"Molecular cloning of cDNA coding for human prepro-urokinase.";			
RT	Gene 36:183-188(1985).			
RN				
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=85203359; PubMed=3888571;			
RA	Jacobs P., Cravador A., Lorian R., Brockly F., Colau B., Chuchana P.,			
RA	van Eise A., Herzog A., Bollen A.;			

RT "Molecular cloning, sequencing, and expression in *Escherichia coli* of  
RT human preprourikinase cDNA." ;  
RT DNA 4:139-146(1985).  
RN [5]  
RP SEQUENCE FROM N.A.  
RA Rieder M.J., Carrington D.P., Chung M.-W., Lee K.L., Peol C.L., Yi Q.,  
RA Nickerson D.A.;  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muny D.M., Sodegren B.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Buterfield V.S.N., Krzywinski M.I., Skalska U., Smilins D.E.,  
RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences." ;  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [7]  
RP SEQUENCE OF 66-431 FROM N.A.  
RX MEDLINE=84272706; PubMed=6589620;  
RA Verde P., Stoppelli M.P., Galeffi P., di Nocera P., Biasi F.;  
RT "Identification and primary sequence of an unspliced human urokinase  
RT poly(A) + RNA." ;  
RL Proc. Natl. Acad. Sci. U.S.A. 81:4727-4731(1984).  
RN [8]  
RP SEQUENCE OF 21-177.  
RX MEDLINE=83055084; PubMed=6754569;  
RA Gunzler W.A., Steffens G.J., Otting F., Kim S.-M.A., Frankus E.,  
RA Flohe L.;  
RT "The primary structure of high molecular mass urokinase from human  
RT urine. The complete amino acid sequence of the A chain." ;  
RL Hoppe-Seyler's Z. Physiol. Chem. 363:1155-1165(1982).  
RN [9]  
RP SEQUENCE OF 156-176 AND 179-224.  
RX MEDLINE=83003608; PubMed=6749491;  
RA Schaller U., Nick H., Rickli E.B., Gillesen D., Leger W.,  
RA Studer R.O.;  
RT "Human low-molecular-weight urinary urokinase. Partial  
RT characterization and preliminary sequence data of the two polypeptide  
RT chains." ;  
RL Eur. J. Biochem. 125:251-257(1982).  
RN [10]  
RP SEQUENCE OF 158-410.  
RX MEDLINE=83055099; PubMed=6754572;  
RA Steffens G.J., Gunzler W.A., Otting F., Frankus E., Flohe L.;  
RT "The complete amino acid sequence of low molecular mass urokinase  
RT from human urine." ;  
RL Hoppe-Seyler's Z. Physiol. Chem. 363:1043-1058(1982).  
RN [11]  
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
RX MEDLINE=96000858; PubMed=8591045;  
RA Sprengon G., Phillips C., Nowak U.K., Ponting C.P., Saunders D.,  
RA Dobson C.M., Stuart D.I., Jones E.Y.;  
RT "The crystal structure of the catalytic domain of human  
RT urokinase-type plasminogen activator." ;  
RL Structure 3:681-691(1995).  
RN [12]  
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 159-411.  
RN [13]  
RP MEDLINE=20266327; PubMed=10805774;  
RA Speit S., Jacob U., Arroyo de Prada N., Sturzebecher J., Wilhelm O.G.,  
RA Bode W., Magdolen V., Huber R., Moroder L.;  
RT "4-(4-aminomethyl)phenylguanidine derivatives as nonpeptidic highly  
RT selective inhibitors of human urokinase." ;  
RL Proc. Natl. Acad. Sci. U.S.A. 97:5113-5118(2000).  
RN [14]  
RP STRUCTURE BY NMR.  
RX MEDLINE=89127526; PubMed=2536903;  
RA Oswald R.E., Bogusky M.J., Bamberger M., Smith R.A.G., Dobson C.M.;  
RT "Dynamics of the multidomain fibrinolytic protein urokinase from two-  
RT dimensional NMR." ;  
RL Nature 337:579-582(1989).  
RN [15]  
RP STRUCTURE BY NMR OF 67-155.  
RX MEDLINE=93003110; PubMed=1337118;  
RA Li X., Smith R.A.G., Dobson C.M.;  
RT "Sequential 1H NMR assignments and secondary structure of the kringle  
RT domain from urokinase." ;  
RL Biochemistry 31:9562-9571(1992).  
RN [16]  
RP STRUCTURE BY NMR OF 67-155.  
RX MEDLINE=94149701; PubMed=8107091;  
RA Li X., Bokman A.M., Llinas M., Smith R.A.G., Dobson C.M.;  
RT "Solution structure of the kringle domain from urokinase-type  
RT plasminogen activator." ;  
RL J. Mol. Biol. 235:1548-1559(1994).  
RN [17]  
RP VARIANT LEU-141.  
RX MEDLINE=96186279; PubMed=8652631;  
RA Yoshimoto M., Ushiyama Y., Sakai M., Tamaki S., Hara H., Takahashi K.,  
RA Sawasaki Y., Hanada K.;  
RT "Characterization of single chain urokinase-type plasminogen  
RT activator with a novel amino-acid substitution in the kringle  
RT structure." ;  
RL Biochim. Biophys. Acta 1293:83-89(1996).  
RN [18]  
RP VARIANT LEU-141.  
RX MEDLINE=97218551; PubMed=9065988;  
RA Come B., Berczy M., Belin D.;  
RT "Detection of polymorphisms in the human urokinase-type plasminogen  
RT activator gene." ;  
RL Thromb. Haemost. 77:434-435(1997).  
RN [19]  
RP ERRATUM.  
RA Come B., Berczy M., Belin D.;  
RL Thromb. Haemost. 78:973-973(1997).  
RN [20]  
RP VARIANT LEU-141.  
RX MEDLINE=97337920; PubMed=9194591;  
RA Turkmen B., Schmitt M., Schmaldt B., Trommler P., Hell W.,  
RA Creutzburg S., Graeff H., Magdolen V.;  
RT "Mutational analysis of the genes encoding urokinase-type plasminogen  
RT activator (uPA) and its inhibitor PAI-1 in advanced ovarian cancer." ;  
RL Electrophoresis 18:686-689(1997).  
RN [21]  
RP FUNCTION: Potent plasminogen activator and is clinically used for  
RP therapy of thrombolytic disorders.  
CC -1- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in  
CC plasminogen to form plasmin.  
CC -1- SUBUNIT: Found in high and low molecular mass forms. Each consists  
CC of two chains, A and B. The high molecular mass form contains a  
CC long chain A. Cleavage occurs after residue 155 in the low  
CC molecular mass form to yield a short A1 chain.  
CC -1- PHARMACEUTICAL: Available under the name Abbockinase (Abbott). Used  
CC in Pulmonary Embolism (PE) to initiate fibrinolysis.  
CC -1- SIMILARITY: Belongs to peptidase family 11.  
CC -1- SIMILARITY: Contains 1 EGF-like domain.  
CC -1- SIMILARITY: Contains 1 kringle domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X02419; CAA26268.1; -
DR EMBL; M15476; AAA61253.1; -
DR EMBL; D00244; BAA00175.1; -
DR EMBL; D11143; BAA01919.1; -
DR EMBL; X02760; CAA26535.1; -
DR EMBL; AF377330; AAK53822.1; -
DR EMBL; BC013575; AAH13575.1; -
DR EMBL; K03226; AAC97138.1; -
DR EMBL; K02286; AAA61252.1; -
DR EMBL; A21571; CAA01559.1; -
DR EMBL; A18397; CAA01390.1; -
DR PIR; A00931; UKHU.
DR PDB; 1KDU; 31-OCT-93.

Query Match      100.0%; Score 44; DB 1; Length 431;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 KPSSPPEE 8
Db      156 KPSSPPEE 163

RESULT 2
UROK_PAPCY
ID UROK_PAPCY STANDARD; PRT; 433 AA.
AC P16227;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Urokinase-type plasminogen activator precursor (BC 3.4.21.73) (uPA)
GN (U-plasminogen activator).
OS PAU.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Papio.
OC NCBI_TaxID=9556;
OX [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Thoracic aorta;
RX MEDLINE=90287734; PubMed=2113276;
RA Au Y.P.T., Wang T.W., Clowes A.W.;
RT "Nucleotide and deduced amino acid sequences of baboon urokinase-type
RT plasminogen activator."
RL Nucleic Acids Res. 18:3411-3411(1990).
CC -1- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in
CC plasminogen to form plasmin.
CC -1- SUBUNIT: Found in high and low molecular mass forms. Each consists
CC of two chains, A and B. The high molecular mass form contains a
CC long chain A. Cleavage occurs after residue 155 in the low
CC molecular mass form to yield a short A1 chain (By similarity).
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 1 EGF-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X51935; CAA36200.1; -
DR PIR; S14687; UKBAY.
DR HSSP; P00749; 1LMW.
DR MEROPS; S01.231; -
DR InterPro; IPR009003; Cys_Ser_tryptsin.

```

```

DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR006001; Kringle.
DR InterPro; IPR008293; Pept_S1a uPA.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00051; Kringle; 1.
DR Pfam; PF00089; trypsin; 1.
DR PIRSF; PIRSF001144; Uro_klaem_act; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00018; KRINGLE.
DR PRODOM; PD000395; Kringle; 1.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00130; KR; 1.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; FALSE_NEG.
DR PROSITE; PSS0026; EGF_3; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS50070; KRINGLE_2; 1.
DR PROSITE; PS50240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
DR KX plasminogen activation; Hydrolyase; Serine protease; Glycoprotein;
KW Kringle; EGF-like domain; Zymogen; Signal.
FT SIGNAL 1 20
FT CHAIN 21 433 UROKINASE-TYPE PLASMINOGEN ACTIVATOR.
FT CHAIN 21 176 CHAIN A (BY SIMILARITY).
FT CHAIN 155 176 SHORT A CHAIN (A1) (BY SIMILARITY).
FT CHAIN 178 433 CHAIN B (BY SIMILARITY).
FT DOMAIN 26 62 EGF-LIKE.
FT DOMAIN 69 150 KRINGLE.
FT DOMAIN 151 177 CONNECTING PEPTIDE.
FT DOMAIN 178 433 SERINE PROTEASE.
FT DISULFID 30 38 BY SIMILARITY.
FT DISULFID 32 50 BY SIMILARITY.
FT DISULFID 52 61 BY SIMILARITY.
FT DISULFID 167 298 INTERCHAIN (BY SIMILARITY).
FT DISULFID 208 224 BY SIMILARITY.
FT DISULFID 216 287 BY SIMILARITY.
FT DISULFID 315 384 BY SIMILARITY.
FT DISULFID 347 363 BY SIMILARITY.
FT DISULFID 374 402 BY SIMILARITY.
FT ACT_SITE 223 223 CHARGE RELAY SYSTEM.
FT ACT_SITE 274 274 CHARGE RELAY SYSTEM.
FT ACT_SITE 378 378 CHARGE RELAY SYSTEM.
FT CARBOHYD 324 324 N-LINKED (GLCNAC...) (BY SIMILARITY).
SQ SEQUENCE 433 AA; 48595 MW; 816D22DFEDC8792 CRC64;

Query Match      100.0%; Score 44; DB 1; Length 433;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 KPSSPPEE 8
Db      155 KPSSPPEE 162

RESULT 3
PEDF_HUMAN
ID PEFH_HUMAN STANDARD; PRT; 418 AA.
AC P36955; Q96CT1; Q96R01; Q9BMA4;
DT 01-JUN-1994 (Rel. 29, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Pigment epithelium-derived factor precursor (PEDF) (EPC-1).
GN SERPINF1 OR PEDF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

```



RC TISSUE=EYE;  
 RA MEDLINE=93165728; PubMed=8434014;  
 RA Steele F.R., Chader G.J., Johnson L.V., Tombran-Tink J.;  
 RT "pigment epithelium-derived factor: neurotrophic activity and  
 RT identification as a member of the serine protease inhibitor gene  
 RT family";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1526-1530(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Yin B., Peng X., Yuan J., Qiang B.;  
 RL Submitted (Jul-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Muscle;  
 RA MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ueda T.B., Tomihyuki S., Carninci P., Frange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
 RA Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butcherfield V.S.N., Krzywinski M.I., Skalska U., Smolins D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [4]  
 RP SEQUENCE OF 1-332 FROM N.A.  
 RC TISSUE=Lung fibroblast;  
 RA Coljee V.W.;  
 RL Theiss (1996), Medical College of Pennsylvania / Philadelphia, U.S.A.  
 RN [5]  
 RP SEQUENCE OF 72-418 FROM N.A.  
 RC TISSUE=Fibroblast;  
 RA MEDLINE=93232057; PubMed=8473338;  
 RA Pignolo R.U., Cristofalo V.U., Rotenberg M.O.;  
 RT "Senescent WI-38 cells fail to express Bpc-1, a gene induced in young  
 RT cells upon entry into the G0 state.";  
 RL J. Biol. Chem. 268:8949-8957(1993).  
 RN [6]  
 RP CHARACTERIZATION.  
 RA MEDLINE=94043087; PubMed=8226833;  
 RA Becerra S.P., Palmer I., Kumar A., Steele F.R., Shiloach J.,  
 RA Notario V., Chader G.J.;  
 RT "Overexpression of fetal human pigment epithelium-derived factor in  
 RT Escherichia coli. A functionally active neurotrophic factor.";  
 RL J. Biol. Chem. 268:23148-23156(1993).  
 RN [7]  
 RP CHARACTERIZATION.  
 RA MEDLINE=96029704; PubMed=7592790;  
 RA Becerra S.P., Sagasti A., Spinelletti P., Notario V.;  
 RT "Pigment epithelium-derived factor behaves like a noninhibitory  
 RT serpin. Neurotrophic activity does not require the serpin reactive  
 RT loop";  
 RL J. Biol. Chem. 270:25992-25999(1995).  
 RN [8]  
 RP X-RAY CRYSTALLOGRAPHY (2.85 ANGSTROMS).  
 RA MEDLINE=21457291; PubMed=11562499;  
 RA Simonovic M., Gettings P.G.W., Volz K.;  
 RT "Crystal structure of human PEDF, a potent anti-angiogenic and neurite  
 RT growth-promoting factor";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:11131-11135(2001).  
 RN [9]  
 RP VARIANT THR-72.

RX MEDLINE=99329357; PubMed=10398730;  
 RA Koeneke R., Pina A.L., Loyer M., Davidson J., Robitaille J.,  
 RA Maunehée I., Tombran-Tink J.;  
 RT "Four polymorphic variations in the PEDF gene identified during the  
 RT mutation screening of patients with Leber congenital amaurosis";  
 RL Mol. Vision 5:10-10(1999).  
 RN [1]  
 RP FUNCTION: NEUTROPHILIC PROTEIN; INDUCES EXTENSIVE NEURONAL  
 CC DIFFERENTIATION IN RETINOBLASTOMA CELLS. POTENT INHIBITOR OF  
 CC ANGIOGENESIS. AS IT DOES NOT UNDERGO THE S (STRESSED) TO R  
 CC (RELAXED) CONFORMATIONAL TRANSITION CHARACTERISTIC OF ACTIVE  
 CC SERPINS, IT EXHIBITS NO SERINE PROTEASE INHIBITORY ACTIVITY.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: RETINAL PIGMENT EPITHELIAL CELLS.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN QUIESCENT CELLS.  
 CC -1- DOMAIN: THE N-TERMINAL (AA 44-121) EXHIBITS NEURITE OUTGROWTH-  
 CC INDUCING ACTIVITY. THE C-TERMINAL EXPOSED LOOP (AA 382-418)  
 CC IS ESSENTIAL FOR SERPIN ACTIVITY.  
 CC -1- SIMILARITY: Belongs to the serpin family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: W76979; AAA60088.1; -;  
 DR EMBL: AF400442; AAK92491.1; -;  
 DR EMBL: BC000522; AAH00522.1; -;  
 DR EMBL: BC013984; AAH13984.1; -;  
 DR EMBL: U57450; AAB38685.1; -;  
 DR EMBL: U57445; AAB38685.1; JOINED.  
 DR EMBL: U57446; AAB38685.1; JOINED.  
 DR EMBL: U57447; AAB38685.1; JOINED.  
 DR EMBL: U57448; AAB38685.1; JOINED.  
 DR EMBL: U57449; AAB38685.1; JOINED.  
 DR EMBL: M90439; AAA93524.1; -;  
 DR PIR: A47281; A47281.  
 DR PDB: 1HW, 03-OCT-01.  
 DR Genew; HGNC:8824; SERPINF1.  
 DR MIM; 172860; -;  
 DR GO: GO:0004867; F:serine protease inhibitor activity; TAS.  
 DR GO: GO:0004868; F:serpin; TAS.  
 DR GO: GO:0008283; P:cell proliferation; TAS.  
 DR GO: GO:0007275; P:development; TAS.  
 DR GO: GO:0007397; P:histogenesis and organogenesis; TAS.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin; 1.  
 DR SMART: SM00093; SERPIN; 1.  
 DR PROSITE: PS00284; SERPIN; 1.  
 KW Serpin; Glycoprotein; Signal; Polymorphism; 3D-structure.  
 FT SIGNAL 1 15  
 FT CHAIN 16 418  
 FT ACT SITE 382 382  
 FT CARBOHYD 285 285  
 FT VARIANT 72 72  
 FT CONFLICT 97 98  
 FT CONFLICT 132 132  
 FT HELIX 45 48  
 FT HELIX 50 72  
 FT TURN 74 75  
 FT STRAND 78 80  
 FT HELIX 82 92  
 FT HELIX 93 95  
 FT HELIX 98 107  
 FT TURN 108 109  
 FT HELIX 110 112  
 FT HELIX 118 129  
 FT TURN 130 130  
 FT TURN 132 133  
 FT STRAND 136 144  
 EQ -> DE (IN REF. 1 AND 4).  
 P -> R (IN REF. 3; AAH13984).  
 /FTID=VAR\_009126.  
 M -> T (in dbSNP:1136287).  
 EQ -> DE (IN REF. 1 AND 4).  
 P -> R (IN REF. 3; AAH13984).

FT TURN 146 147  
 FT HELIX 152 162  
 FT STRAND 167 168  
 FT HELIX 173 187  
 FT TURN 188 190  
 FT STRAND 205 214  
 FT STRAND 217 219  
 FT HELIX 223 225  
 FT STRAND 227 232  
 FT STRAND 238 256  
 FT TURN 257 260  
 FT STRAND 261 268  
 FT TURN 269 271  
 FT STRAND 272 279  
 FT HELIX 287 290  
 FT TURN 291 292  
 FT HELIX 295 304  
 FT STRAND 306 315  
 FT STRAND 317 324  
 FT HELIX 326 330  
 FT TURN 331 335  
 FT HELIX 336 339  
 FT TURN 344 346  
 FT STRAND 353 364  
 FT STRAND 368 370  
 FT STRAND 387 389  
 FT STRAND 394 400  
 FT TURN 401 403  
 FT STRAND 406 412  
 FT TURN 415 416  
 SQ SEQUENCE 418 AA, 46342 MW, 29B573A62EAS1BBS CRC64;

Query Match 81.8%; Score 36; DB 1; Length 418;  
 Best Local Similarity 85.7%; Pred. No. 72;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PSSPPEE 8  
 Db 22 PASPPEE 28

RESULT 4  
 YB96\_HUMAN STANDARD; PRT; 892 AA.  
 AC Q96KM6; Q9ULM4; Rel. 41, Created  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical zinc finger protein KIAA1196.  
 GN KIAA1196.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21638749; Pubmed=11780052;  
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.W., Brown A.J.,  
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,  
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,  
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 RA Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Lehesvallo M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McCormack L.J., McLay K., McMurtry A.A.,  
 RA Mite S.A., Mistry D., Moore M.U.F., Mullikin J.C., Nickerson T.,

RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 RA Phillimore B.J.C.T., Prachalingam S.R., Plumb R.W., Ramsay H.,  
 RA Rice C.M., Rose M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,  
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Suleston J.E.,  
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.;  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871(2001).  
 RN [2]  
 RP SEQUENCE OF 42-892 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=20039619; Pubmed=10574462;  
 RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,  
 RA Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XV.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 6:337-345(1999).  
 RN [3]  
 RP SEQUENCE OF 425-892 FROM N.A.  
 RC TISSUE=Testis;  
 RA Duesterhoeft A., Lauber J., Mewes H.-W., Weil B., Wiemann S.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: May function as a transcription factor.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
 CC FINGER PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AL118506; CAC15498.3; -;  
 DR EMBL; AB033022; BAA86510.1; -;  
 DR EMBL; AL834525; CAD39181.1; -;  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF00096; zf-C2H2; 5.  
 DR SMART; SM00355; Znf\_C2H2; 6.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 4.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 4.  
 KW Hypothetical protein; Transcription regulation; DNA-binding;  
 KW Metal-binding; Zinc-finger; Nuclear protein; Repeat.  
 FT ZN\_FING 105 129 C2H2-TYPE (ATYPICAL).  
 FT ZN\_FING 140 163 C2H2-TYPE.  
 FT ZN\_FING 540 563 C2H2-TYPE.  
 FT ZN\_FING 594 618 C2H2-TYPE (ATYPICAL).  
 FT ZN\_FING 630 653 C2H2-TYPE (ATYPICAL).  
 FT ZN\_FING 750 774 C2H2-TYPE (ATYPICAL).  
 FT ZN\_FING 784 807 C2H2-TYPE.  
 SQ SEQUENCE 892 AA; 97264 MW; 50449C476DFCE4DF CRC64;

Query Match 81.8%; Score 36; DB 1; Length 892;  
 Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PSSPPEE 8  
 Db 407 PASPPEE 413

RESULT 5  
 PIGO\_HUMAN STANDARD; PRT; 1088 AA.  
 AC O8TEQ8; O8TDS8; Q96CS9; Q9Y4B0;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)

15-MAR-2004 (Rel. 43, last annotation update)  
 Phosphatidylinositol-glycan biosynthesis, class O protein (PIG-O).  
 Homo sapiens (Human).  
 CC Takayoga; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Testis;  
 RA Ansoyge W., Witkner U., Mewes H.-W., Gassenhuber J., Wiemann S.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Spleen;  
 RA Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T., Ohara O.;  
 RT "The nucleotide sequence of a long cDNA clone isolated from human  
 spleen";  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RA Lameudin J.E., McCready P.M., Skowronski E., Adamson A.W.,  
 RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Scilwagen S.,  
 RA Phan H., Velasco N., Ganes J., Dangman L., Poundstone P.,  
 RA Christensen M., Georgescu A., Avila J., Liu S., Attix C., Andreise T.,  
 RA Tranheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,  
 RA Bruce R., Thomas P., Quan G., Krommiller B., Arellano A.,  
 RA Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.O.,  
 RA Carraro A.V.;  
 RT "Sequence analysis of a human P1 clone containing the XRCC3 DNA repair  
 gene";  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC TISSUE=Ductenium;  
 RA MEDLINE=2388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Shier G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.T., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Lonnellano N.A., Peters G.J., Abramson R.D., Millaby S.J.,  
 RA Boeak S.A., McMan P.J., McKernan K.J., Malek J.A., Gumaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Ketterman M., Madan A.C., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smilins D.E.,  
 RA Scheraga A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [5]  
 RP SEQUENCE OF 601-968 FROM N.A.  
 RA Takeda S., Kadowaki S., Haga T., Takeasu H., Mitaku S.;  
 RT "Identification of G protein-coupled receptor genes from the human  
 genome sequence";  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Involved in GPI-anchor biosynthesis but not essential  
 for this process. Involved, together with PIGF, in the transfer of  
 ethanolamine phosphate to the third mannose of GPI (By  
 similarity).  
 CC -1- PATHWAY: GPI-anchor biosynthesis.  
 CC -1- SUBUNIT: Forms a complex with PIGF. PIGF is required to stabilize  
 PIGO (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic  
 reticulum (By similarity).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;

CC Comment=Additional isoforms seem to exist;  
 CC Name=1;  
 CC IsoId=Q8TEQ8-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q8TEQ8-2; Sequence=VSP\_003944;  
 CC Note=No experimental confirmation available;  
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to a  
 frameshift in position 46.  
 CC -1- CAUTION: Ref.2 sequence differs from that shown due to  
 frameshifts.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AL833956; CAD38806.1; ALT\_FRAME.  
 DR EMBL: AK074064; BAB84890.1; ALT\_FRAME.  
 DR EMBL: AK090433; BAC03414.1; ALT\_FRAME.  
 DR EMBL: AC004472; AAC07985.1; ALT\_SEQ.  
 DR EMBL: BC029271; AAH29271.1; -.  
 DR EMBL: BC001030; AAH01030.1; ALT\_INIT.  
 DR EMBL: BC013987; AAH13987.1; ALT\_INIT.  
 DR EMBL: AB083625; BAB89338.1; -.  
 DR Genew; HGNC:23215; PIGO.  
 DR GO; GO:0005789; C:Endoplasmic reticulum membrane; ISS.  
 DR GO; GO:0006506; P:GPI anchor biosynthesis; ISS.  
 DR InterPro; IPR002591; Phosphodiect.  
 DR Pfam; PF01663; Phosphodiect; 1.  
 DR GPI-anchor biosynthesis; Transmembrane; Endoplasmic reticulum;  
 KW Alternative splicing.  
 FT TRANSMEM 4  
 FT TRANSMEM 456 476 POTENTIAL.  
 FT TRANSMEM 481 501 POTENTIAL.  
 FT TRANSMEM 509 529 POTENTIAL.  
 FT TRANSMEM 540 560 POTENTIAL.  
 FT TRANSMEM 574 594 POTENTIAL.  
 FT TRANSMEM 667 687 POTENTIAL.  
 FT TRANSMEM 700 720 POTENTIAL.  
 FT TRANSMEM 746 766 POTENTIAL.  
 FT TRANSMEM 829 849 POTENTIAL.  
 FT TRANSMEM 856 876 POTENTIAL.  
 FT TRANSMEM 943 963 POTENTIAL.  
 FT TRANSMEM 1013 1033 POTENTIAL.  
 FT TRANSMEM 1047 1067 POTENTIAL.  
 FT VARSPLIC 448 864 Missing (in isoform 2).  
 FT TRANSMEM 179 181 DIT -> ARG (IN REF. 4; AAH13987).  
 FT TRANSMEM 225 231 DIVIAHF -> EVSNQHV (IN REF. 4; AAH01030).  
 FT TRANSMEM 350 350 G -> Y (IN REF. 4; AAH29271).  
 FT TRANSMEM 353 353 D -> Y (IN REF. 4; AAH29271).  
 FT TRANSMEM 415 415 R -> KG (IN REF. 4).  
 FT TRANSMEM 883 968 PFTVQWASAVWMTATQFVSGHQPVPVPAIHMAHAFVGF  
 PEHGSCTWLPALVAVANTPASHLLPRAVGCPLLLPPLPCE  
 SQGL -> KYLSDSLKNSDVSAAPVFEVLLPLFLSLT  
 EGMPTTRKVFVLSILPAIAKQIDPSCWFPQFMERRDKS  
 SSTKPCGNAASS (IN REF. 5).  
 FT TRANSMEM 1088 118669 MW; 065985AC16D1D340 CRC64;  
 SQ SEQUENCE 1088 AA; 118669 MW;  
 Query Match 81.8%; Score 36; DB 1; Length 1088;  
 Best Local Similarity 85.7%; Pred. No. 2e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Oy 2 PSSPPER 8  
 Db 308 PSTPPER 314  
 RESULT 6  
 SIRA\_YEAST

ID SIR4\_YEAST STANDARD; PRT; 1358 AA.  
AC P11978;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, last sequence update)  
DT 28-FEB-2003 (Rel. 41, last annotation update)  
DE Regulatory protein SIR4 (silent information regulator 4).  
GN SIR4 OR STS OR ASD1 OR UTH2 OR YDR227W OR YD934.12.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88142836; Pubmed=3325825;  
RA Marshall M., Mahoney D., Rose A., Hicks J.B., Broach J.R.;  
RT "Functional domains of SIR4, a gene required for position effect  
RT regulation in Saccharomyces cerevisiae.";  
RL Mol. Cell. Biol. 7:4441-4452(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95192063; Pubmed=7885847;  
RA Davies C.J., Hutchison C.A. III;  
RT "Insertion site specificity of the transposon Tn3.";  
RL Nucleic Acids Res. 23:507-514(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX STRAIN=6288c / AB972;  
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A.;  
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP REVIEW.  
RX MEDLINE=21579783; Pubmed=11722841;  
RA Gasser S.M., Cockell M.M.;  
RT "The molecular biology of the SIR proteins.";  
RL Gene 279:1-16(2001).  
CC -1- FUNCTION: The proteins SIR1 through SIR4 are required for  
CC transcriptional repression of the silent mating type loci, HML and  
CC HMR. The proteins SIR2 through SIR4 repress multiple loci by  
CC modulating chromatin structure. Involves the compaction of  
CC chromatin fiber into a more condensed form.  
CC -1- SUBUNIT: Homodimer and interacts with SIR1, SIR2, SIR3 and RAP1 C-  
CC terminus.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M37249; AAA20881.1; -  
CC EMBL; U13239; AAC3144.1; -  
CC EMBL; Z48612; CAA88507.1; -  
CC PIR; A29360; A29360.  
CC GerMOnline; 140719; -  
CC SGD; S0002635; SIR4.  
CC GO; GO:0000783; C:nuclear telomere cap complex; IDA.  
CC GO; GO:0005724; C:nuclear telomeric heterochromatin; IDA.  
CC GO; GO:0006303; P:double-strand break repair via nonhomologous recombination; IDA.  
CC Transcription regulation; Repressor; DNA-binding; Nuclear protein;  
CC Coiled coil.  
CC VARIANT 1277 1347 COILED COIL (POTENTIAL).  
CC P -> L.  
CC FT DOMAIN 994 994  
CC FT VARIANT 994 994  
CC SO SEQUENCE 1358 AA; 152061 MW; 96C98765964F094E CRC64;  
Query Match 81.8%; Score 36; DB 1; Length 1358;  
Best Local Similarity 85.7%; Pred. No. 2.Se+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1131 KPSSPPD 1137  
RESULT 7  
ID CA1B\_HUMAN STANDARD; PRT; 1806 AA.  
AC P12107; Q14034; Q9UIT4; Q9UIT5; Q9UIT6;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 28-FEB-2003 (Rel. 41, last sequence update)  
DT 10-OCT-2003 (Rel. 42, last annotation update)  
DE Collagen alpha 1(XI) chain precursor.  
GN COL11A1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90202924; Pubmed=1690726;  
RA Yoshioke H., Ramirez F.;  
RT "Pro-alpha 1(XI) collagen. Structure of the amino-terminal propeptide  
RT and expression of the gene in tumor cell lines.";  
RL J. Biol. Chem. 265:6423-6426(1990).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORMS A, B AND C), AND VARIANTS STL2/MARSHALL  
RP SYNDROME ARG-676; 921-GLN--PRO-926 DEL; 1313-PHE--GLY-1315 DEL AND  
RP VAL-1516.  
RX MEDLINE=20455728; Pubmed=10463316;  
RA Annunen S., Koerkoe J., Czarny M., Warman M.L., Brunner H.G.,  
RA Kaeferlaenen H., Mulliken J.B., Tranebjaerg L., Brooks D.G.,  
RA Cox G.P., Cruysberg J.R., Curtis M.A., Davenport S.L.H.,  
RA Friedrich C.A., Kallila I., Krawczynski M.R., Latos-Bielenska A.,  
RA Mkal S., Olsen B.R., Shimo N., Somer M., Vakkula M., Zlotogora J.,  
RA Prockop D.J., Ala-Kokko L.;  
RT "Splicing mutations of 54-bp exons in the COL11A1 gene cause Marshall  
RT syndrome, but other mutations cause overlapping Marshall/Stickler  
RT phenotypes.";  
RL Am. J. Hum. Genet. 65:974-983(1999).  
RN [3]  
RP SEQUENCE OF 538-1806 FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=89034222; Pubmed=1182841;  
RA Bernard M., Yoshioke H., Rodriguez E., van der Rest M., Kimura T.,  
RA Ninomiya Y., Olsen B.R., Ramirez F.;  
RT "Cloning and sequencing of pro-alpha 1(XI) collagen cDNA  
RT demonstrates that type XI belongs to the fibrillar class of collagens  
RT and reveals that the expression of the gene is not restricted to  
RT cartilaginous tissue.";  
RL J. Biol. Chem. 263:17159-17166(1988).  
RN [4]  
RP ALTERNATIVE SPLICING.  
RC TISSUE=Blood;  
RX MEDLINE=95238468; Pubmed=7721876;  
RA Zhidkova N.I., Justice S.K., Mayne R.;  
RT "Alternative mRNA processing occurs in the variable region of the  
RT pro-alpha 1(XI) and pro-alpha 2(XI) collagen chains.";  
RL J. Biol. Chem. 270:9486-9493(1995).  
RN [5]  
RP VARIANT STL2 VAL-625.  
RX MEDLINE=97026296; Pubmed=8872475;  
RA Richards A.J., Yates J.R.W., Williams R., Payne S.J., Pope F.M.,  
RA Scott J.D., Sneed M.P.;  
RT "A family with Stickler syndrome type 2 has a mutation in the COL11A1  
RT gene resulting in the substitution of glycine 97 by valine in  
RT alpha-1(XI) collagen.";  
RL Hum. Mol. Genet. 5:1339-1343(1996).  
CC -1- FUNCTION: May play an important role in fibrillogenesis by  
CC controlling lateral growth of collagen II fibrils.  
CC -1- SUBUNIT: Trimers composed of three different chains: alpha 1(XI),  
CC alpha 2(XI), and alpha 3(XI). Alpha 3(XI) is a post-translational  
CC modification of alpha 1(XI). Alpha 1(XI) can also be found instead  
CC of alpha 3(XI)=1(XI).  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=3;

CC Comment=Additional isoforms seem to exist. There is alternative  
CC usage of exon IIA or exon IIB. Transcripts containing exon IIA  
CC or IIB are present in cartilage, but exon IIB is preferentially  
CC utilized in transcripts from tendon;  
CC Name=A;  
CC IsoId=PI2107-1; Sequence=Displayed;  
CC Name=B;  
CC IsoId=PI2107-2; Sequence=VSP\_001145;  
CC Name=C;  
CC IsoId=PI2107-3; Sequence=VSP\_001146;  
CC -1- TISSUE SPECIFICITY: Cartilage, placenta and some tumor or virally  
CC transformed cell lines. Isoforms using exon IIA or IIB are found  
CC in the cartilage while isoforms using only exon IIB are found in  
CC the tendon.  
CC -1- PFM: Proline at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -1- DISEASE: Defects in COL1A1 are the cause of Stickler syndrome  
CC type 2 (STL2) [MIM:604841]; also known as Stickler syndrome  
CC vitreous type 2, or beaded vitreous type, due to the presence of  
CC irregularly thickened fiber bundles throughout vitreous cavity.  
CC Stickler syndrome (hereditary progressive arthro-ophthalmopathy)  
CC is an autosomal dominant disorder characterized by progressive  
CC myopia beginning in the first decade of life, vitreo-retinal  
CC degeneration, retinal detachment, cleft palate, midfacial  
CC hypoplasia, osteoarthritis, and sensorineural hearing loss.  
CC -1- DISEASE: Defects in COL1A1 are the cause of Marshall syndrome  
CC [MIM:154780], an autosomal dominant disorder with ocular, oro-  
CC facial, auditory and skeletal manifestations. It shares several  
CC features with Stickler syndrome, such as midfacial hypoplasia,  
CC high myopia, and sensorineural-hearing deficit.  
CC -1- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.  
CC -1- SIMILARITY: HIGH, TO ALPHA 1(V) AND ALPHA 3(V) CHAINS.  
CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
DR EMBL, J04177; AAA51891.1; -;  
DR EMBL, AF101112; AAF04724.1; JOINED.  
DR EMBL, AF101079; AAF04724.1; JOINED.  
DR EMBL, AF101080; AAF04724.1; JOINED.  
DR EMBL, AF101081; AAF04724.1; JOINED.  
DR EMBL, AF101082; AAF04724.1; JOINED.  
DR EMBL, AF101083; AAF04724.1; JOINED.  
DR EMBL, AF101084; AAF04724.1; JOINED.  
DR EMBL, AF101085; AAF04724.1; JOINED.  
DR EMBL, AF101086; AAF04724.1; JOINED.  
DR EMBL, AF101087; AAF04724.1; JOINED.  
DR EMBL, AF101088; AAF04724.1; JOINED.  
DR EMBL, AF101089; AAF04724.1; JOINED.  
DR EMBL, AF101090; AAF04724.1; JOINED.  
DR EMBL, AF101091; AAF04724.1; JOINED.  
DR EMBL, AF101092; AAF04724.1; JOINED.  
DR EMBL, AF101093; AAF04724.1; JOINED.  
DR EMBL, AF101094; AAF04724.1; JOINED.  
DR EMBL, AF101095; AAF04724.1; JOINED.  
DR EMBL, AF101096; AAF04724.1; JOINED.  
DR EMBL, AF101097; AAF04724.1; JOINED.  
DR EMBL, AF101098; AAF04724.1; JOINED.  
DR EMBL, AF101099; AAF04724.1; JOINED.  
DR EMBL, AF101100; AAF04724.1; JOINED.  
DR EMBL, AF101101; AAF04724.1; JOINED.  
DR EMBL, AF101102; AAF04724.1; JOINED.  
DR EMBL, AF101103; AAF04724.1; JOINED.  
DR EMBL, AF101104; AAF04724.1; JOINED.  
DR EMBL, AF101105; AAF04724.1; JOINED.  
DR EMBL, AF101106; AAF04724.1; JOINED.  
DR EMBL, AF101107; AAF04724.1; JOINED.

DR	EMBL	AF1011108	AAE047224.1	JOINED
DR	EMBL	AF1011109	AAE047224.1	JOINED
DR	EMBL	AF1011110	AAE047224.1	JOINED
DR	EMBL	AF1011111	AAE047225.1	JOINED
DR	EMBL	AF1011112	AAE047225.1	JOINED
DR	EMBL	AF1010709	AAE047225.1	JOINED
DR	EMBL	AF1010870	AAE047225.1	JOINED
DR	EMBL	AF1010881	AAE047225.1	JOINED
DR	EMBL	AF1010882	AAE047225.1	JOINED
DR	EMBL	AF1010883	AAE04725.1	JOINED
DR	EMBL	AF1010884	AAE04725.1	JOINED
DR	EMBL	AF1010885	AAE04725.1	JOINED
DR	EMBL	AF1010886	AAE04725.1	JOINED
DR	EMBL	AF1010887	AAE04725.1	JOINED
DR	EMBL	AF1010888	AAE04725.1	JOINED
DR	EMBL	AF1010889	AAE04725.1	JOINED
DR	EMBL	AF1010890	AAE04725.1	JOINED
DR	EMBL	AF101091	AAE04725.1	JOINED
DR	EMBL	AF101092	AAE04725.1	JOINED
DR	EMBL	AF101093	AAE04725.1	JOINED
DR	EMBL	AF101094	AAE04725.1	JOINED
DR	EMBL	AF101095	AAE04725.1	JOINED
DR	EMBL	AF101096	AAE04725.1	JOINED
DR	EMBL	AF101097	AAE04725.1	JOINED
DR	EMBL	AF101098	AAE04725.1	JOINED
DR	EMBL	AF101099	AAE04725.1	JOINED
DR	EMBL	AF101100	AAE04725.1	JOINED
DR	EMBL	AF101101	AAE04725.1	JOINED
DR	EMBL	AF101102	AAE04725.1	JOINED
DR	EMBL	AF101103	AAE04725.1	JOINED
DR	EMBL	AF101104	AAE04725.1	JOINED
DR	EMBL	AF101105	AAE04725.1	JOINED
DR	EMBL	AF101106	AAE04725.1	JOINED
DR	EMBL	AF101107	AAE04725.1	JOINED
DR	EMBL	AF101108	AAE04725.1	JOINED
DR	EMBL	AF101109	AAE04725.1	JOINED
DR	EMBL	AF101110	AAE04725.1	JOINED
DR	EMBL	AF101111	AAE04725.1	JOINED
DR	EMBL	AF101112	AAE04726.1	JOINED
DR	EMBL	AF101117	-	JOINED
DR	EMBL	AF1010109	AAE04726.1	JOINED
DR	EMBL	AF1010801	AAE04726.1	JOINED
DR	EMBL	AF1010802	AAE04726.1	JOINED
DR	EMBL	AF1010803	AAE04726.1	JOINED
DR	EMBL	AF1010804	AAE04726.1	JOINED
DR	EMBL	AF1010805	AAE04726.1	JOINED
DR	EMBL	AF1010806	AAE04726.1	JOINED
DR	EMBL	AF1010807	AAE04726.1	JOINED
DR	EMBL	AF1010808	AAE04726.1	JOINED
DR	EMBL	AF1010809	AAE04726.1	JOINED
DR	EMBL	AF1010810	AAE04726.1	JOINED
DR	EMBL	AF101091	AAE04726.1	JOINED
DR	EMBL	AF101092	AAE04726.1	JOINED
DR	EMBL	AF101093	AAE04726.1	JOINED
DR	EMBL	AF101094	AAE04726.1	JOINED
DR	EMBL	AF101095	AAE04726.1	JOINED
DR	EMBL	AF101096	AAE04726.1	JOINED
DR	EMBL	AF101097	AAE04726.1	JOINED
DR	EMBL	AF101098	AAE04726.1	JOINED
DR	EMBL	AF101099	AAE04726.1	JOINED

```

Query Match          81.8%; Score 36; DB 1; Length 1806;
Best Local Similarity 75.0%; Pred. No. 3.4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy      1 KPSSPPEE 8
        ||:|||||
Db       392 KPITSPNPN 399

RESULT 8
TAP2_NPVAC
ID_TAP2_NPVAC
AC_P41454; STANDARD; PRT; 249 AA.
```

```

DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable apoptosis inhibitor 2 (IAP-2).
GN IAP2.
OS Autographa californica nuclear polyhedrosis virus (AcMNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolydnavirus.
OX NCBI_TaxID=46015;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C6;
RX MEDLINE=94303173; PubMed=8030224;
RA Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;
RT "The complete DNA sequence of Autographa californica nuclear
RT polyhedrosis virus."
RL Virology 202:586-605(1994).
CC -1- SIMILARITY: Contains 1 BIR repeat.
CC -1- SIMILARITY: Contains 1 RING-type zinc finger.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L23858; AAC66701.1; -.
DR PIR: H72858; H72858.
DR InterPro: IPR001370; BIR.
DR InterPro: IPR001841; Znf_ring.
DR Pfam: PF00653; BIR; 1.
DR SMART: SM00238; BIR; 1.
DR SMART: SM00184; RING; 1.
DR PROSITE: PS01282; BIR_REPEAT_1; FALSE_NEG.
DR PROSITE: PS50143; BIR_REPEAT_2; 1.
DR PROSITE: PS00518; ZF_RING_1; FALSE_NEG.
DR PROSITE: PS50089; ZF_RING_2; 1.
DR KW Apoptosis; Zinc-finger.
FT REPEAT 85 152 BIR.
FT ZN_FING 202 237 RING-TYPE.
SQ SEQUENCE 249 AA; 28621 MW; 4DE2825A792BE520 CRC64;
OY 1 KPSSPPEE 8
Db 186 KPSPAPAE 193
Query Match 79.5%; Score 35; DB 1; Length 249;
Best Local Similarity 75.0%; Pred. No. 61;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

RL J. Biol. Chem. 271:18954-18960(1996).
CC -1- FUNCTION: Fibronectin bind cell surfaces and various compounds
CC including collagen, fibrin, heparin, DNA, and actin. Fibronectins
CC are involved in cell adhesion, cell motility, opsonization, wound
CC healing, and maintenance of cell shape.
CC -1- FUNCTION: Isoform 2 is probably involved in matrix organization of
CC cartilage.
CC -1- SUBUNIT: Mostly heterodimers or multimers of alternatively spliced
CC variants, connected by 2 disulfide bonds near the carboxyl ends;
CC to a lesser extent homodimers.
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Each of the "extra domain" and the connecting strand 3
CC are present in some forms of fibronectin and absent from others;
CC Name=1;
CC IsoId=Q28275-1; Sequence=Displayed;
CC Name=2; Synonyms=(V+C)-;
CC IsoId=Q28275-2; Sequence=VSP_003251, VSP_003252;
CC Note=lacks repeat 15 of fibronectin type-III, repeat 10 of
CC fibronectin type-I, and the connecting strand 3;
CC -1- TISSUE SPECIFICITY: Plasma FN (soluble dimeric form) is secreted
CC by hepatocytes. Cellular FN (dimeric or cross-linked multimeric
CC form), made by fibroblasts, epithelial and other cell types, is
CC deposited as fibrils in the extracellular matrix. Isoform 2 is the
CC major transcript in articular cartilage, but it is absent from
CC liver.
CC -1- PTM: Sulfated (By similarity).
CC -1- SIMILARITY: Contains at least 4 fibronectin type III domains.
CC -1- SIMILARITY: Contains at least 3 fibronectin type I domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U52106; AAC48612.1; -.
DR EMBL: U52105; AAC48611.1; -.
DR HSP: P02751; 1PNI.
DR InterPro: IPR000085; Fibnctn1.
DR InterPro: IPR008957; FN_III-1like.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003962; FNIII_subd.
DR Pfam: PF00039; fn1; 3.
DR Pfam: PF00041; fn3; 3.
DR PRINTS: PR00014; FNTYPEIII.
DR SMART: SM00058; FN1; 2.
DR SMART: SM00060; FN3; 3.
DR PROSITE: PS01253; FIBRONECTIN_1; 2.
DR KW Glycoprotein; Plasma; Heparin-binding; Acute phase; Cell adhesion;
KW Repeat; Sulfation; Alternative splicing.
FT NON_TER 1
FT DOMAIN <1 204 HEPARIN-BINDING 2 (BY SIMILARITY).
FT DOMAIN 325 >522 FIBRIN-BINDING 2 (BY SIMILARITY).
FT DOMAIN <1 25 FIBRONECTIN TYPE-III 13.
FT DOMAIN 26 114 FIBRONECTIN TYPE-III 14.
FT DOMAIN 115 204 FIBRONECTIN TYPE-III 15.
FT DOMAIN 205 315 CONNECTING STRAND 3 (CS-3) (V REGION).
FT DOMAIN 316 395 FIBRONECTIN TYPE-III 16.
FT DOMAIN 417 461 FIBRONECTIN TYPE-I 10.
FT DOMAIN 462 504 FIBRONECTIN TYPE-I 11.
FT DOMAIN 506 >522 FIBRONECTIN TYPE-I 12.
FT SITE 304 306 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 419 448 BY SIMILARITY.
FT DISULFID 446 458 BY SIMILARITY.
FT DISULFID 464 491 BY SIMILARITY.
FT DISULFID 489 501 BY SIMILARITY.
FT DISULFID 508 >522 BY SIMILARITY.
FT MOD_RES 509 509 SULFATION (POTENTIAL).
FT CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (BY SIMILARITY).

```

```

CC CARBOHYD 277 277 O-LINKED (GALNAC. . .) (BY SIMILARITY).
CC CARBOHYD 278 278 O-LINKED (GALNAC. . .) (BY SIMILARITY).
CC VARSPPLIC 205 205 D -> E (in isoform 2).
CC VARSPPLIC 206 462 /FTid=VSP_003251.
CC VARSPPLIC 206 462 Missing (in isoform 2).
CC VARSPPLIC 206 462 /FTid=VSP_003252.
CC NON TER 522 522
CC SEQUENCE 522 AA; 57700 MW; DB1D9A54C2BD4E26 CRC64;

Query Match 79.5%; Score 35; DB 1; Length 522;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8
Db 152 KPSSPPEE 159

RESULT 10
F1NC HORSE STANDARD; PRT; 522 AA.
AC 028377; 028378;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Fibronectin (FN) (Fragment).
GN FNI.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Liver;
RX MEDLINE=96324983; PubMed=8702559;
RA MacLeod J.N., Burton-Wurster N., Gu D.N., Lust G.;
RT "Fibronectin mRNA splice variant in articular cartilage lacks bases
RT encoding the V, III-15, and I-10 protein segments."
RL J. Biol. Chem. 271:18954-18960(1996).
CC -1- FUNCTION: Fibronectin bind cell surfaces and various compounds
CC including collagen, fibrin, heparin, DNA, and actin. Fibronectins
CC are involved in cell adhesion, cell motility, opsonization, wound
CC healing, and maintenance of cell shape.
CC -1- FUNCTION: Isoform 2 is probably involved in matrix organization of
CC cartilage.
CC -1- SUBUNIT: MOSTLY HETERODIMERS OR MULTIMERS OF ALTERNATIVELY SPLICED
CC VARIANTS, CONNECTED BY 2 DISULFIDE BONDS NEAR THE CARBOXYL ENDS;
CC TO A LESSER EXTENT HOMODIMERS.
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist. Each of the "extra
CC domain" and the connecting strand 3 are present in some forms of
CC fibronectin and absent in others;
CC Name=1;
CC IsoId=Q28377-1; Sequence=Displayed;
CC Name=2; Synonyms=(V+C)-;
CC IsoId=Q28377-2; Sequence=VSP_003253; VSP_003254;
CC Note=lacks repeat 15 of fibronectin type-III, repeat 10 of
CC fibronectin type-I, and the connecting strand 3;
CC -1- TISSUE SPECIFICITY: Plasma FN (soluble dimeric form) is secreted
CC by hepatocytes. Cellular FN (dimeric or cross-linked multimeric
CC forms), made by fibroblasts, epithelial and other cell types, is
CC deposited as fibrils in the extracellular matrix. Isoform 2 is the
CC major transcript in articular cartilage; very low levels in lymph
CC node, bone, aorta, and skin; absent from liver, spleen, placenta,
CC cardiac muscle, skeletal muscle, stomach, small intestine, and
CC kidney.
CC -1- PTM: Sulfated (By similarity).
CC -1- SIMILARITY: Contains at least 4 fibronectin type III domains.
CC -1- SIMILARITY: Contains at least 3 fibronectin type I domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U52107; AAC48613.1; -
CC EMBL; U52108; AAC48614.1; -
CC HSSP; P02751; FNI.
CC InterPro; IPR000083; Fibnctn1.
CC InterPro; IPR008957; FN III-like.
CC InterPro; IPR003961; FN III.
CC InterPro; IPR003962; FNIII_subd.
CC Pfam; PF00039; fn1_3.
CC Pfam; PF00041; fn3_3.
CC PRINTS; PR00014; FNTYPEIII.
CC SMART; SM00058; FNI; 2.
CC SMART; SM00060; FN3; 3.
CC PROSITE; PS01253; FIBRONECTIN 1; 2.
CC Glycoprotein; Plasma; Heparin-binding; Acute phase; Cell adhesion;
CC Repeat; Sulfation; Alternative splicing.
CC Repeat; Sulfation; Alternative splicing.
CC NON TER 1 1
CC DOMAIN 1 204 HEPARIN-BINDING 2 (BY SIMILARITY).
CC DOMAIN 325 >522 FIBRIN-BINDING 2 (BY SIMILARITY).
CC DOMAIN 1 25 FIBRONECTIN TYPE-III 13.
CC DOMAIN 26 114 FIBRONECTIN TYPE-III 14.
CC DOMAIN 115 204 FIBRONECTIN TYPE-III 15.
CC DOMAIN 205 315 CONNECTING STRAND 3 (CS-3) (V REGION).
CC DOMAIN 316 395 FIBRONECTIN TYPE-III 16.
CC DOMAIN 417 461 FIBRONECTIN TYPE-I 10.
CC DOMAIN 462 504 FIBRONECTIN TYPE-I 11.
CC DOMAIN 506 >522 FIBRONECTIN TYPE-I 12.
CC SITE 304 306 CELL ATTACHMENT SITE (POTENTIAL).
CC SITE 419 408 BY SIMILARITY.
CC DISULFID 446 458 BY SIMILARITY.
CC DISULFID 464 491 BY SIMILARITY.
CC DISULFID 489 501 BY SIMILARITY.
CC MOD RES 515 515 SUPRATON (POTENTIAL).
CC CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
CC CARBOHYD 277 277 O-LINKED (GALNAC. . .) (BY SIMILARITY).
CC CARBOHYD 278 278 O-LINKED (GALNAC. . .) (BY SIMILARITY).
CC VARSPPLIC 205 205 D -> E (in isoform 2).
CC VARSPPLIC 206 462 /FTid=VSP_003253.
CC VARSPPLIC 206 462 Missing (in isoform 2).
CC VARSPPLIC 206 462 /FTid=VSP_003254.
CC NON TER 522 522
CC SEQUENCE 522 AA; 57577 MW; 893E8AC895864D41 CRC64;

Query Match 79.5%; Score 35; DB 1; Length 522;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8
Db 152 KPSSPPEE 159

RESULT 11
ID ASPH BOVIN STANDARD; PRT; 754 AA.
AC 028056;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Aspartyl/asparaginyl beta-hydroxylase (EC 1.14.11.16) (Aspartate beta-
DE hydroxylase) (ASP beta-hydroxylase) (peptide-aspartate beta-
DE dioxygenase).
GN ASPH.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.

```





entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL, D25328; BAA04998.1; -.

DR EMBL, BC002536; AAH02536.1; -.

DR EMBL, BC029138; AAH29138.1; -.

DR EMBL, M64784; AAA36435.1; -.

DR PIR, JC2055; JC2055.

DR HSSP, P00512; 3PFK.

DR Genew, HGNC:8878; PFKP.

DR GK, Q01813; -.

DR MIM, 171840; -.

DR GO, GO:0005945; C:6-phosphofructokinase complex; NAS.

DR GO, GO:0003872; P:6-phosphofructokinase activity; TAS.

DR GO, GO:0006096; P:glycolysis; NAS.

DR InterPro, IPR000023; Pfructkinase.

DR Pfam, PF00365; PFK; 2.

DR PIRSF, PIRSF000533; PFK euk; 1.

DR PRINTS, PR00476; PPFCTKINASE.

DR ProDom, PD000707; Ppfuckinase; 2.

DR PROSITE, PS00433; PHOSPHOFRUCTOKINASE; 2.

DR KINASE, TRANSFERASE; Glycolysis; Repeat; Allosteric enzyme; Phosphorylation; Magnesium; Multigene family.

FT CONFLICT 484 485 PG -> IP (IN REF. 3).

FT CONFLICT 498 498 MISSING (IN REF. 3).

FT CONFLICT 699 699 A -> B (IN REF. 3).

SO SEQUENCE 784 AA; 85596 MW; 22522E7E9A9F0F6 CRC64;

Query Match 79.5%; Score 35; DB 1; Length 784; Best Local Similarity 85.7%; Pred. No. 2.1e+02; Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 PSSPPE 8

Db 240 PSSPPE 246

RESULT 13

K6PP\_MOUSE

ID K6PP\_MOUSE STANDARD; PRT; 784 AA.

AC Q9WUJ3; Q9UI86;.

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE 6-phosphofructokinase, type C (EC 2.7.1.11) (Phosphofructokinase 1) (Phosphohexokinase) (Phosphofructo-1-kinase isozyme C) (PFK-C).

GN PFKP OR PFKC.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10990;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Brain;

RA Gunasekera D., Kemp R.G.;

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Sanchez-Martinez C., Estevez A.M., Argon J.J.;

RT "Cloning, expression and properties of the phosphofructokinase C isozyme from ascites tumor cells."

RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6;

RA Gunasekera D., Kemp R.G.;

RT "Genomic organization, 5' flanking region and tissue-specific expression of mouse phosphofructokinase C gene."

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=FVB/N-3; TISSUE=Mammary gland;

RX MEDLINE=23388257; PubMed=1247932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Krausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.P., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ussid T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Muliyil S.J., Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smolins D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -1- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-fructose 1,6-bisphosphate.

CC -1- COFACTOR: Magnesium (By similarity).

CC -1- ENZYME REGULATION: Allosteric enzyme activated by ADP, AMP, or fructose bisphosphate and inhibited by ATP or citrate (By similarity).

CC -1- PATHWAY: Key control step of glycolysis.

CC -1- SIMILARITY: Belongs to the phosphofructokinase family. Two domains subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL, AF123533; AAD23571.1; -.

DR EMBL, Y19008; CAB64347.1; -.

DR EMBL, AF249893; AAF75700.1; -.

DR EMBL, AF250369; AAF75700.1; JOINED.

DR EMBL, AF250370; AAF75700.1; JOINED.

DR EMBL, AF250371; AAF75700.1; JOINED.

DR EMBL, AF250372; AAF75700.1; JOINED.

DR EMBL, AF251021; AAF75700.1; JOINED.

DR EMBL, BC006926; AAH06926.1; -.

DR HSSP, P00512; 3PFK.

DR MGD, MGI:1891833; PFKP.

DR GO, GO:0003872; P:6-phosphofructokinase activity; IDA.

DR InterPro, IPR000023; Ppfuckinase.

DR Pfam, PF00365; PFK; 2.

DR PIRSF, PIRSF000533; PFK euk; 1.

DR PRINTS, PR00476; PPFCTKINASE.

DR ProDom, PD000707; Ppfuckinase; 2.

DR PROSITE, PS00433; PHOSPHOFRUCTOKINASE; 2.

DE KINASE, TRANSFERASE; Glycolysis; Repeat; Allosteric enzyme; Phosphorylation; Magnesium; Multigene family.

FT CONFLICT 180 180 T -> P (IN REF. 3).

SO SEQUENCE 784 AA; 85455 MW; E9C5AABF26FCA65 CRC64;

Query Match 79.5%; Score 35; DB 1; Length 784; Best Local Similarity 85.7%; Pred. No. 2.1e+02; Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 PSSPPE 8

Db 239 PSSPPE 245

RESULT 14

K6PP\_RAT

ID K6PP\_RAT STANDARD; PRT; 786 AA.

```

AC P47860;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 6-phosphofructokinase, type C (EC 2.7.1.11) (phosphofructokinase
DE 1) (phosphohexokinase) (phosphofructo-1-kinase isozyme C) (PFK-C)
DE (fragment).
GN PFKP OR PFKC OR PFK-C.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Hypothalamus;
RX MEDLINE=94148828; PubMed=8106374;
RA Gekker N., Johnson R.C., Jenkins A., Mains R.E., Sul H.S.;
RT "Structure, distribution, and functional expression of the
RT phosphofructokinase C isozyme.";
RL J. Biol. Chem. 269:3348-3355(1994).
RN [2]
RP IDENTIFICATION OF PROBABLE FRAMESHIFT.
RA Bairoch A.;
RL Unpublished observations (DEC-1995).
CC -1- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-
CC fructose 1,6-bisphosphate.
CC -1- COFACTOR: Magnesium.
CC -1- ENZYME REGULATION: Allosteric enzyme activated by ADP, AMP, or
CC fructose bisphosphate and inhibited by ATP or citrate.
CC -1- PATHWAY: Key control step of glycolysis.
CC -1- TISSUE SPECIFICITY: Expressed at high level in neuroendocrine
CC tissues.
CC -1- SIMILARITY: Belongs to the phosphofructokinase family. Two domains
CC subfamily.
CC -1- CAUTION: This is a conceptual translation; a frameshift was
CC introduced in position 25 to increase the similarity in the
CC N-terminal region with orthologs.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L25387; AAA17757.1; ALT_FRAME.
DR HSSP: P00512; 3PFK.
DR InterPro: IPR000023; Pffructkinase.
DR Pfam: PF00365; PFK; 2.
DR PIRSF: PIRSF000533; PFK euk; 1.
DR PRINTS: PR00476; PFRCTKINASE.
DR ProDom: PD000707; Pfructkinase; 2.
DR PROSITE, PS00433; PHOSPHOFRUCTOKINASE; 2.
DR Kinase; Transferase; Glycolysis; Repeat; Allosteric enzyme;
KW Phosphorylation; Magnesium; Multigene family.
FT NON TER 1 1
SQ SEQUENCE 786 AA; 85601 MW; 9EE17A20A8F3B077 CRC64;

Query Match 79.5%; Score 35; DB 1; Length 786;
Best Local Similarity 85.7%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PSSPPEE 8
Db 238 PESPPPE 244

RESULT 15
K6PP_RABIT STANDARD; PRT; 791 AA.
AC P47859;
DT 01-FEB-1996 (Rel. 33, Created)

```

```

DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 6-phosphofructokinase, type C (EC 2.7.1.11) (phosphofructokinase
DE 1) (phosphohexokinase) (phosphofructo-1-kinase isozyme C) (PFK-C).
GN PFKP.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=94164929; PubMed=8119919;
RA Li Y., Valaitis A.P., Latshaw S.P., Kwiatkowska D., Tripathi R.L.,
RA Campbell M.C., Kemp R.G.;
RT "Structure and expression of the cDNA for the C isozyme of
RT phosphofructo-1-kinase from rabbit brain.";
RL J. Biol. Chem. 269:5781-5787(1994).
CC -1- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-
CC fructose 1,6-bisphosphate.
CC -1- COFACTOR: Magnesium.
CC -1- ENZYME REGULATION: Allosteric enzyme activated by ADP, AMP, or
CC fructose bisphosphate and inhibited by ATP or citrate.
CC -1- PATHWAY: Key control step of glycolysis.
CC -1- SIMILARITY: Belongs to the phosphofructokinase family. Two domains
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U01154; AAA17707.1; -.
DR HSSP: P00512; 3PFK.
DR InterPro: IPR000023; Pffructkinase.
DR Pfam: PF00365; PFK; 2.
DR PIRSF: PIRSF000533; PFK euk; 1.
DR PRINTS: PR00476; PFRCTKINASE.
DR ProDom: PD000707; Pfructkinase; 2.
DR PROSITE, PS00433; PHOSPHOFRUCTOKINASE; 2.
DR Kinase; Transferase; Glycolysis; Repeat; Allosteric enzyme;
KW Phosphorylation; Magnesium; Multigene family.
SQ SEQUENCE 791 AA; 86349 MW; 3C10A36F229FDBE8 CRC64;

Query Match 79.5%; Score 35; DB 1; Length 791;
Best Local Similarity 85.7%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PSSPPEE 8
Db 240 PESPPPE 246

RESULT 16
DPOL_HSV21 STANDARD; PRT; 1240 AA.
ID DPOL_HSV21
AC P07938;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase (EC 2.7.7.7).
OS Herpes simplex virus (type 2 / strain 186).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10312;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87277385; PubMed=3038677;
RA Tsurumi T., Maeno K., Nishiyama Y.;
RT "Nucleotide sequence of the DNA polymerase gene of herpes simplex

```

```

RT Virus type 2 and comparison with the type 1 counterpart.;
RL Gene 52:129-137(1987).
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + (DNA)(N).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M16321; AAA45853.1; -.
CC PIR: A27315; DUBB21.
CC InterPro: IPR006172; DNA_pol_B.
CC InterPro: IPR006134; DNA_pol_B_dom.
CC InterPro: IPR006133; DNA_pol_B_exo.
CC Pfam: PF00136; DNA_pol_B_1.
CC Pfam: PF03104; DNA_pol_B_exo; 1.
CC PRINTS: PR00106; DNAPOLB.
CC SMART: SM00486; POLBC; 1.
CC PROSITE: PS00116; DNA_POLYMERASE_B; 1.
CC Transferrase; DNA-directed DNA polymerase; DNA replication;
CC DNA-binding; Nuclear protein.
CC KW SEQUENCE 1240 AA; 13735 MW; 1CDABD1D1F8180E CRC64;

Query Match 79.5%; Score 35; DB 1; Length 1240;
Best Local Similarity 85.7%; Pred. No. 3.4e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 PSSPPEE 8
Db 798 PSSPPEE 804

RESULT 17
TABP_HUMAN STANDARD; PRT; 1729 AA.
AC Q9C0C2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 182 kDa tankyrase 1-binding protein.
GN TNKS1BP1 OR TAB182 OR KIAA1741.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21082932; PubMed=11214970;
RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:347-355(2000).
RN (3)
RN SEQUENCE OF 495-1729 FROM N.A.
RP TISSUE=Spleen;
RC MEDLINE=22579292; PubMed=12693554;

```

```

RA Jikuya H., Takano J., Kikuno R., Hirosewa M., Nagase T., Nomura N.,
RA Ohara O.;
RT "Characterization of long cDNA clones from human adult spleen. II. The
RT complete sequences of 81 cDNA clones.";
RL DNA Res. 10:49-57(2003).
CC -1- SUBUNIT: Binds to the ANK repeat domain of TNKS1 and TNKS2.
CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic. Colocalizes with
CC chromosomes during mitosis, and in the cytoplasm with cortical
CC actin.
CC -1- TISSUE SPECIFICITY: Detected in testis, ovary, lung, skeletal
CC muscle, heart, prostate and pancreas, and at very low levels in
CC brain and peripheral blood leukocytes.
CC -1- PTM: ADP-ribosylated by TNKS1 (in vitro).
CC -1- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC in position 1071, 1097 and 1467.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF441771; AAM1553.1; -.
CC EMBL: AB051528; BAB21832.2; ALT_INT.
CC EMBL: AK074113; BAB84939.1; ALT_FRAME.
CC Genew; HGNC:19081; TNKS1BP1.
CC GK; Q9C0C2; -.
CC DR MIM; 607104; -.
CC DR GO; GO:0005737; Cytoplasm; NAS.
CC DR GO; GO:0005724; Cytosolic telomeric heterochromatin; NAS.
CC DR GO; GO:0005634; Cytosol; NAS.
CC DR GO; GO:000506; F-actin binding; NAS.
CC DR GO; GO:0019899; Enzyme binding; NAS.
CC DR GO; GO:007004; P-kinase-dependent telomere maintenance; NAS.
CC DR InterPro: IPR008979; Gal bind like.
CC KW Nuclear protein; Chromosomal protein; ADP-ribosylation.
CC FT DOMAIN 2 103 ARG/GHU/LYS/PRO-RICH (CHARGED).
CC FT DOMAIN 127 767 PRO-RICH.
CC FT DOMAIN 210 1572 ACIDIC.
CC FT DOMAIN 1010 1340 GLY-RICH.
CC FT DOMAIN 1450 1542 TANKYRASE-BINDING.
CC FT DOMAIN 1572 1729 ARG/GHU/LYS-RICH (CHARGED).
CC FT DOMAIN 1629 1635 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT DOMAIN 1723 1729 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT CONFLICT 84 84 P -> L (IN REF. 2).
CC FT CONFLICT 322 322 S -> T (IN REF. 2).
CC FT CONFLICT 388 388 S -> P (IN REF. 2).
CC FT CONFLICT 554 554 Q -> H (IN REF. 3).
CC FT CONFLICT 604 604 P -> S (IN REF. 1).
CC FT CONFLICT 1450 1450 F -> S (IN REF. 2).
CC SEQUENCE 1729 AA; 181814 MW; C65F38FA37045C4A CRC64;

Query Match 79.5%; Score 35; DB 1; Length 1729;
Best Local Similarity 75.0%; Pred. No. 4.8e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8
Db 366 KPSSPPEE 373

RESULT 18
BAZA_HUMAN STANDARD; PRT; 1878 AA.
AC Q9UIF9; O00536; O15030; Q96H26;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Bromodomain adjacent to zinc finger domain 2A (Transcription
DE termination factor-1 interacting protein 5) (TF-I interacting protein
DE 5) (Tifs) (hMALp3).

```

GN BAZZA OR TIPS OR KIA0314.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NC NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=20130112; PubMed=10662543;  
 RA Jones M.H., Hamana N., Nezu J.-I., Shimane M.;  
 RT "A novel family of bromodomain genes.";  
 RL Genomics 63:40-45(2000).  
 RN [2]  
 RP SEQUENCE OF 332-738 FROM N.A.  
 RC TISSUE=Brain;  
 RA Jansa P., Grunnet I.;  
 RL Submitted (Apr-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 639-1878 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=97349984; PubMed=9205841;  
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,  
 RT Prediction of the coding sequences of unidentified human genes. VII.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.";  
 RL DNA Res. 4:141-150(1997).  
 RN [4]  
 RP SEQUENCE OF 1035-1878 FROM N.A.  
 RC TISSUE=Lymph;  
 RX MEDLINE=42388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heich F.,  
 RA Datchenko L., Marais K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Scapleton M., Soares W.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshitsuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W.,  
 RA Viallano D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Matra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- FUNCTION: May play a role in transcriptional regulation  
 CC interacting with ISM1. May serve a specific role in maintaining or  
 CC altering the chromatin structure of the rDNA locus (By  
 CC similarity).  
 CC -1- SUBUNIT: Together with ISM1/SNP2h, it forms a complex termed NORC  
 CC (nucleolar remodeling complex).  
 CC -1- SUBCELLULAR LOCATION: Nuclear. Colocalizes with the basal RNA  
 CC polymerase I transcription factor UBF in the nucleolus.  
 CC -1- TISSUE SPECIFICITY: Expressed at moderate levels in most tissues  
 CC analyzed, including heart, brain, placenta, lung, skeletal muscle,  
 CC kidney and pancreas.  
 CC -1- SIMILARITY: Belongs to the MAL family.  
 CC -1- SIMILARITY: Contains 4 A.T hook DNA-binding repeats.  
 CC -1- SIMILARITY: Contains 1 bromodomain.  
 CC -1- SIMILARITY: Contains 1 DDT domain.  
 CC -1- SIMILARITY: Contains 1 methyl-binding (MBD) domain.  
 CC -1- SIMILARITY: Contains 1 PHD-type zinc finger.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AB032254; BAA89211.1; -;  
 CC EMBL: AF000422; AAB60864.1; -;  
 CC EMBL: AB002312; BAA20773.1; -;  
 CC EMBL: BC008965; AAB08965.1; -;  
 CC HSSP: Q92831; IB91.  
 CC Genew: HGNC:962; BAZ2A.  
 CC MIM: 605682; -;  
 CC GO: GO:0005731; C:nucleolus organizer complex; NAS.  
 CC GO: GO:0003677; F:DNA binding; NAS.  
 CC GO: GO:0030528; F:transcription regulator activity; NAS.  
 CC GO: GO:0006338; P:chromatin modeling; NAS.  
 CC GO: GO:0006355; P:regulation of transcription, DNA-dependent; NAS.  
 CC InterPro: IPR000637; AT hook.  
 CC InterPro: IPR001487; Bromodomain.  
 CC InterPro: IPR004022; DDT\_dom.  
 CC InterPro: IPR001739; Methyl-Cpg\_bind.  
 CC InterPro: IPR001965; Znf PHD.  
 CC Pfam: PF02178; AT hook; 4.  
 CC Pfam: PF00439; bromodomain; 1.  
 CC Pfam: PF02791; DDT; 1.  
 CC Pfam: PF01429; MBD; 1.  
 CC Pfam: PF00628; PHD; 1.  
 CC PRINTS: PR00503; BROMODOMAIN.  
 CC SMART: SM00384; AT hook; 4.  
 CC SMART: SM00297; BROMO; 1.  
 CC SMART: SM00571; DDT; 1.  
 CC SMART: SM00391; MBD; 1.  
 CC SMART: SM00249; PHD; 1.  
 CC PROSITE: PS00633; BROMODOMAIN\_1; FALSE\_NEG.  
 CC PROSITE: PS50014; BROMODOMAIN\_2; 1.  
 CC PROSITE: PS50827; DDT; 1.  
 CC PROSITE: PS01359; ZF\_PHD\_1; FALSE\_NEG.  
 CC PROSITE: PS50016; ZF\_PHD\_2; 1.  
 CC PROSITE: PS50016; ZF\_PHD\_2; 1.  
 CC Nuclear protein; Repeat; DNA-binding.  
 CC DOMAIN 525 537  
 CC FT DNA\_BIND 622 634 A.T HOOK 1.  
 CC FT DNA\_BIND 643 655 A.T HOOK 2.  
 CC FT DOMAIN 821 886 DDT.  
 CC FT DNA\_BIND 1159 1171 A.T HOOK 3.  
 CC FT DNA\_BIND 1377 1389 A.T HOOK 4.  
 CC FT ZN\_FING 1649 1699 PHD-TYPE.  
 CC FT DOMAIN 1783 1853 BROMODOMAIN.  
 CC FT DOMAIN 633 772 LYS-RICH.  
 CC FT DOMAIN 666 765 COILED COIL (POTENTIAL).  
 CC FT DOMAIN 1185 1250 GLU-RICH.  
 CC FT DOMAIN 1263 1384 PRO-RICH.  
 CC FT DOMAIN 1732 1735 POLY-ARG.  
 CC FT DOMAIN 574 700 L->V (IN REF. 2).  
 CC FT CONFLICT 700 700 L->Q (IN REF. 2 AND 3).  
 CC FT CONFLICT 720 720 H->Q (IN REF. 2 AND 3).  
 CC FT CONFLICT 727 738 SKAKKKKKTKQ->KIRKKKKKKKK (IN REF.  
 CC 2).  
 CC FT CONFLICT 785 785 K->R (IN REF. 3).  
 CC FT CONFLICT 951 951 P->L (IN REF. 3).  
 CC FT CONFLICT 1005 1006 GR->EG (IN REF. 3).  
 CC FT CONFLICT 1035 1037 IAA->GIR (IN REF. 4).  
 CC FT CONFLICT 1163 1163 G->S (IN REF. 1).  
 CC FT CONFLICT 1166 1166 R->L (IN REF. 1).  
 CC FT CONFLICT 1172 1172 S->F (IN REF. 1).  
 CC FT CONFLICT 1178 1178 L->F (IN REF. 1).  
 CC FT CONFLICT 1202 1202 A->V (IN REF. 1).  
 CC FT CONFLICT 1292 1292 P->L (IN REF. 1).  
 CC FT CONFLICT 1295 1295 L->F (IN REF. 1).  
 CC FT CONFLICT 1313 1313 P->L (IN REF. 1).  
 CC FT CONFLICT 1407 1410 MISSING (IN REF. 4).  
 CC FT CONFLICT 1416 1416 R->P (IN REF. 1).  
 CC FT CONFLICT 1541 1541 R->P (IN REF. 1).  
 CC FT CONFLICT 1571 1571 E->K (IN REF. 1).

```

FT CONFLICT 1616 1616 V -> I (IN REF. 1)
FT CONFLICT 1622 1622 E -> Q (IN REF. 1)
FT CONFLICT 1629 1629 Q -> H (IN REF. 1)
FT CONFLICT 1636 1636 Q -> H (IN REF. 1)
FT CONFLICT 1739 1739 R -> K (IN REF. 1)
FT CONFLICT 1754 1754 G -> R (IN REF. 1)
SQ SEQUENCE 1878 AA; 208480 MW; 400970CA682317 CRC64;

Query Match 79.5%; Score 35; DB 1; Length 1878;
Best Local Similarity 85.7%; Pred. No. 5.2e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 PSSPPE 8
Db 1274 PSSPPE 1280

RESULT 19
PTPD_HUMAN STANDARD; PRT; 1912 AA.
AC P23468;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protein-tyrosine phosphatase delta precursor (BC 3.1.3.48) (R-PTP-
delta).
GN PTPRD.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND MUTAGENESIS OF ARG-1178.
RX MEDLINE=9520468; PubMed=7896816;
RA Pullido R., Krueger N.X., Serra-Pages C., Saito H., Streuli M.;
RT "Molecular characterization of the human transmembrane protein-
RT tyrosine phosphatase delta. Evidence for tissue-specific expression of
RT alternative human transmembrane protein-tyrosine phosphatase delta
RT isoforms."
RT J. Biol. Chem. 270:6722-6728(1995).
RL [2]
RN SEQUENCE OF 390-1912 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=91006018; PubMed=2170109;
RA Krueger N.X., Streuli M., Saito H.;
RT "Structural diversity and evolution of human receptor-like protein
RT tyrosine phosphatases."
RL EMBO J. 9:3241-3252(1990).
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC tyrosine + phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=P23468-1; Sequence=Displayed;
CC Name=2; Synonyms=Kidney;
CC IsoId=P23468-2; Sequence=VSP_005147, VSP_005148, VSP_005149;
CC Name=3; Synonyms=Fetal brain;
CC IsoId=P23468-3; Sequence=VSP_005150;
CC -1- FROM THE TRANSMEMBRANE SEGMENT.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domain.
CC -1- SIMILARITY: Contains 8 fibronectin type III domain.
CC -1- SIMILARITY: Contains 2 protein-tyrosine phosphatase domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL: L38929; AAC41749.1; -
DR EMBL: X54133; CA38068.1; -
DR PIR: A56178; A56178.
DR HSSP: P18052; 1YFO.
DR Genew: HGNC:9668; PTPRD.
DR MIM: 601598; -
DR GO: GO:0005887; C: integral to plasma membrane; TAS.
DR GO: GO:0005001; F: transmembrane receptor protein tyrosine pho. . .; TAS.
DR GO: GO:0006470; P: protein amino acid dephosphorylation; TAS.
DR GO: GO:0007185; P: transmembrane receptor protein tyrosine pho. . .; TAS.
DR InterPro: IPR008957; FN III-1like.
DR InterPro: IPR003961; FN III.
DR InterPro: IPR003962; FNIII subd.
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003598; Ig_c2.
DR InterPro: IPR000387; TYR_phosphatase.
DR Pfam: PF00041; fn3; 8.
DR Pfam: PF00047; ig; 3.
DR Pfam: PF00102; Y_phosphatase; 2.
DR PRINTS: PR00014; ENTPEPITL.
DR SMART: SM00060; FN3; 8.
DR SMART: SM00408; IGc2; 2.
DR SMART: SM00194; PTPC; 2.
DR PROSITE: PS50835; IG_LIKE; 3.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 2.
DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 2.
DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 2.
KM Hydroxylase; Receptor; Glycoprotein; Signal; Transmembrane; Repeat;
KW Immunoglobulin domain; Alternative splicing.
FT SIGNAL 1 20
FT CHAIN 21 1912
FT DOMAIN 21 1265
FT TRANSMEM 1266 1290
FT DOMAIN 1291 1912
FT DOMAIN 126 114
FT DOMAIN 126 224
FT DOMAIN 236 318
FT DOMAIN 320 414
FT DOMAIN 417 513
FT DOMAIN 516 606
FT DOMAIN 609 708
FT DOMAIN 711 822
FT DOMAIN 825 916
FT DOMAIN 918 1017
FT DOMAIN 1020 1137
FT DOMAIN 1375 1618
FT DOMAIN 1619 1912
FT ACT_SITE 1553 1553
FT ACT_SITE 1844 1844
FT SITE 1175 1178
FT CARBOHYD 254 254
FT CARBOHYD 299 299
FT CARBOHYD 724 724
FT CARBOHYD 832 832
FT VARSPLIC 181 189
FT VARSPLIC 226 229
FT VARSPLIC 775 783
FT VARSPLIC 609 1137
FT MUTAGEN 1178 1178
SQ SEQUENCE 1912 AA; 214759 MW; 3A86C8CD32182E26 CRC64;

Query Match 79.5%; Score 35; DB 1; Length 1912;
Best Local Similarity 62.5%; Pred. No. 5.3e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 KPSPPEE 8  
 DB 608 KPSAPPOD 615

RESULT 20  
 ID PINS\_HUMAN STANDARD; PRT; 1948 AA.  
 AC Q13332; Q15718; Q16341;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Receptor-type protein-tyrosine phosphatase 5 precursor (EC 3.1.3.48)  
 DE (R-PTP-S) (Protein-tyrosine phosphatase sigma) (R-PTP-sigma).  
 GN PTPRS.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=96102179; PubMed=8524829;  
 RA Pulido R., Serra-Pages C., Tang M., Streuli M.;  
 RT "The LAR/PTP delta/PTP sigma subfamily of transmembrane protein-  
 RT tyrosine-phosphatases: multiple human LAR, PTP delta, and PTP sigma  
 RT isoforms are expressed in a tissue-specific manner and associate with  
 RT the LAR-interacting protein LIP-1.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:11686-11690 (1995).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=96255038; PubMed=8992885;  
 RA Endo N., Rutledge S.J., Opas E.E., Vogel R., Rodan G.A., Schmidt A.;  
 RT "Human protein tyrosine phosphatase-sigma: alternative splicing and  
 RT inhibition by bisphosphonates.";  
 RL J. Bone Miner. Res. 11:535-543 (1996).  
 RN [3]  
 RN SEQUENCE OF 1-126 FROM N.A.  
 RA Lameirdin J.E., McCreedy P.M., Skowronski E., Viswanathan V.,  
 RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stijlwegen S.,  
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J., Liu S.,  
 RA Dangnan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,  
 RA Attix C., Andreasson T., Frankheim M., Amico-Keller G., Coeffield J.,  
 RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,  
 RA Arellano A., Saunders C., Ow D., Nolan M., Trong S., Kobayashi A.,  
 RA Olsen A.S., Carraro A.V.;  
 RT "Sequence analysis of a 2.5 Mb region in 19p13.3.";  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RN SEQUENCE OF 1503-1589 FROM N.A.  
 RX MEDLINE=92119637; PubMed=1370651;  
 RA Adachi M., Sekiya M., Arimura Y., Takekawa M., Itoh F., Hinoda Y.,  
 RA Imai K., Yachi A.;  
 RT "Protein-tyrosine phosphatase expression in pre-B cell NALM-6.";  
 RL Cancer Res. 52:737-740 (1992).  
 CC -1- FUNCTION: Interacts with LAR-interacting protein LIP-1.  
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein  
 CC tyrosine + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=5;  
 CC Comment=Additional isoforms seem to exist;  
 CC Name=PTPS;  
 CC IsoId=Q13332-1; Sequence=Displayed;  
 CC Name=PTPS-MEA; Sequence=VSP\_050021;  
 CC IsoId=Q13332-2; Sequence=VSP\_050021;  
 CC Name=PTPS-MEB;  
 CC IsoId=Q13332-3; Sequence=VSP\_050022, VSP\_050026, VSP\_050027;  
 CC Name=PTPS-MEC;  
 CC IsoId=Q13332-4; Sequence=VSP\_050024;  
 CC Name=PTPS-P4-7;  
 CC IsoId=Q13332-5; Sequence=VSP\_050023, VSP\_050025;

CC -1- TISSUE SPECIFICITY: Detected in all tissues tested except for  
 CC placenta and liver.  
 CC -1- SIMILARITY: BELONGS TO THE RECEPTOR CLASS OF THE PROTEIN-TYROSINE  
 CC PHOSPHATASE FAMILY.  
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domain.  
 CC -1- SIMILARITY: Contains 8 fibronectin type III domain.  
 CC -1- SIMILARITY: Contains 2 protein-tyrosine phosphatase domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U35234; AAC50299.1; -;  
 DR EMBL; U40317; AAC50567.1; -;  
 DR EMBL; AC005790; AAC62832.1; -;  
 DR EMBL; S78080; AAB21146.2; -;  
 DR HSSP; P18052; IYFO.  
 DR Genew; HGNC:9681; PTPRS.  
 DR MIM; 601576; -;  
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.  
 DR GO; GO:0005001; F:transmembrane receptor protein tyrosine pho. . .; TAS.  
 DR InterPro; IPR008957; FN III-like.  
 DR InterPro; IPR003961; FN III.  
 DR InterPro; IPR003962; FNIII subd.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_C2.  
 DR InterPro; IPR000387; Tyr\_phosphatase.  
 DR InterPro; IPR000242; Tyr\_PP.  
 DR Pfam; PF00041; fn3; 8.  
 DR Pfam; PF00047; Ig; 3.  
 DR Pfam; PF00102; Y\_phosphatase; 2.  
 DR PRINTS; PR00014; ERTYPRITL.  
 DR PRINTS; PR00700; ERTYPRPHASE.  
 DR SMART; SM00060; FN3; 7.  
 DR SMART; SM00408; IGC2; 3.  
 DR SMART; SM00194; PTPC; 2.  
 DR PROSITE; PS50835; IG\_LIKE; 3.  
 DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 2.  
 DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 2.  
 DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 2.  
 DR Hydrolase; Receptor; Glycoprotein; Signal; Transmembrane;  
 KM Cell adhesion; Immunoglobulin domain; Alternative splicing; Repeat.  
 FT FT SIGNAL 1 29  
 FT CHAIN 30 1948  
 FT FT  
 FT DOMAIN 30 1282  
 FT TRANSMEM 1283 1303  
 FT DOMAIN 1304 1948  
 FT DOMAIN 33 123  
 FT DOMAIN 135 233  
 FT DOMAIN 245 327  
 FT DOMAIN 329 423  
 FT DOMAIN 426 522  
 FT DOMAIN 525 615  
 FT DOMAIN 618 717  
 FT DOMAIN 720 831  
 FT DOMAIN 834 926  
 FT DOMAIN 928 1033  
 FT DOMAIN 1036 1151  
 FT DOMAIN 1393 1648  
 FT DOMAIN 1680 1930  
 FT DOMAIN 641 644  
 FT DISULFID 54 107  
 FT DISULFID 156 216  
 FT DISULFID 266 311  
 FT ACT\_SITE 1589 1589  
 FT ACT\_SITE 1880 1880  
 FT ACT\_SITE 1880 1880

POTENTIAL.  
 CYTOPLASMIC (POTENTIAL).  
 IG-LIKE C2-TYPE 1.  
 IG-LIKE C2-TYPE 2.  
 IG-LIKE C2-TYPE 3.  
 FIBRONECTIN TYPE-III 1.  
 FIBRONECTIN TYPE-III 2.  
 FIBRONECTIN TYPE-III 3.  
 FIBRONECTIN TYPE-III 4.  
 FIBRONECTIN TYPE-III 5.  
 FIBRONECTIN TYPE-III 6.  
 FIBRONECTIN TYPE-III 7.  
 FIBRONECTIN TYPE-III 8.  
 PROTEIN-TYROSINE PHOSPHATASE 1.  
 PROTEIN-TYROSINE PHOSPHATASE 2.  
 POLY-PRO.  
 POTENTIAL.  
 POTENTIAL.  
 PHOSPHOCYSTEINE INTERMEDIATE (BY  
 SIMILARITY)  
 PHOSPHOCYSTEINE INTERMEDIATE (BY  
 SIMILARITY).

```

FT CARBOHYD 263 263 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 308 308 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 733 733 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 940 940 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 190 198 Missing (in isoform PTPS-MEA).
FT VARSPLIC 236 239 Missing (in isoform PTPS-MEB).
FT VARSPLIC 617 1034 Missing (in isoform PTPS-F4-7).
FT VARSPLIC 784 792 Missing (in isoform PTPS-MEC).
FT VARSPLIC 1035 1035 Missing (in isoform PTPS-F4-7).
FT VARSPLIC 1350 1365 Missing (in isoform PTPS-MEB).
FT VARSPLIC 1366 1366 Missing (in isoform PTPS-MEB).
FT VARSPLIC 1366 1366 S -> G (in isoform PTPS-MEB).
FT CONFLICT 310 310 /FTID=VSP_050027.
FT CONFLICT 428 429 T -> HP (IN REF. 2).
FT CONFLICT 742 745 LGPV -> RSPA (IN REF. 2).
FT CONFLICT 763 773 GAEGRPPR -> RREGRGRS (IN REF. 2).
FT CONFLICT 910 910 R -> P (IN REF. 2).
FT CONFLICT 986 995 AAEPGAEINAV -> GLSRARRTL (IN REF. 2).
FT CONFLICT 1195 1196 TV -> SL (IN REF. 2).
FT CONFLICT 1431 1431 E -> S (IN REF. 2).
FT CONFLICT 1546 1546 E -> D (IN REF. 4).
FT CONFLICT 1587 1587 V -> A (IN REF. 4).
FT CONFLICT 1705 1705 N -> K (IN REF. 2).
SQ SEQUENCE 1948 AA; 217080 MM; 7DC049ECC03171136 CRC64;

```

```

Query Match 79.5%; Score 35; DB 1; Length 1948;
Best Local Similarity 62.5%; Pred. No. 5.4e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 KSSPPEE 8
Db 617 KPSAPPOD 624

RESULT 21
F1NC BOVIN STANDARD; PRT; 2265 AA.
AC P07589;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Fibronectin (FN).
GN FN1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_Taxid=9913;
RN [1]
RP SEQUENCE.
RX MEDLINE=87054047; PubMed=3780752;
RA Skorstengaard K., Jensen M.S., Sahl P., Petersen T.E., Magnusson S.;
RT "Complete primary structure of bovine plasma fibronectin.";
RL Eur. J. Biochem. 161:441-453(1986).
RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE=83117805; PubMed=6218503;
RA Petersen T.E., Thorgersen H.C., Skorstengaard K., Vibe-Pedersen K.,
RT "Partial primary structure of bovine plasma fibronectin: three types
RT of internal homology.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:137-141(1983).
RN [3]
RP SEQUENCE OF 2170-2265 FROM N.A.
RX MEDLINE=83221567; PubMed=6304699;
RA Kornblith A.R., Vibe-Pedersen K., Baralle F.E.;
RT "Isolation and characterization of cDNA clones for human and bovine

```

```

RT fibronectins.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:3218-3222(1983).
CC -1- FUNCTION: Fibronectins bind cell surfaces and various compounds
CC including collagen, fibrin, heparin, DNA, and actin. Fibronectins
CC are involved in cell adhesion, cell motility, opsonization, wound
CC healing, and maintenance of cell shape.
CC -1- SUBUNIT: MOSTLY HETERODIMERS OR MULTIMERS OF ALTERNATIVELY SPLICED
CC VARIANTS. CONNECTED BY 2 DISULFIDE BONDS NEAR THE CARBOXYL ENDS;
CC TO A LESSER EXTENT HOMODIMERS.
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=1;
CC Comment=A number of isoforms are produced. Each of the "extra
CC domain" and the connecting strand 3 are present in some forms of
CC fibronectin and absent in others;
CC Name=1;
CC IsoId=P07589-1; Sequence=displayed;
CC -1- TISSUE SPECIFICITY: Plasma FN (soluble dimeric form) is secreted
CC by hepatocytes. Cellular FN (dimeric or cross-linked multimeric
CC forms), made by fibroblasts, epithelial and other cell types, is
CC deposited as fibrils in the extracellular matrix.
CC -1- PTM: Sulfated (by similarity).
CC -1- SIMILARITY: Contains 12 fibronectin type I domains.
CC -1- SIMILARITY: Contains 2 fibronectin type II domains.
CC -1- SIMILARITY: Contains 15 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K00800; AAA30521.2; -.
DR PIR; A26452; FNBO.
DR HSSP; P02751; 2FN2.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR000083; Fibnctn1.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR000562; FN_Type_II.
DR InterPro; IPR003962; FnIII_subd.
DR Pfam; PR00039; fn1; 12.
DR Pfam; PR00040; fn2; 2.
DR Pfam; PR00041; fn3; 15.
DR PRINTS; PR00013; FNTYPEII.
DR PRINTS; PR00014; FNTYPEIII.
DR ProDom; PD000995; FN_Type_II; 2.
DR SMART; SM00058; FN1; 12.
DR SMART; SM00059; FN2; 2.
DR SMART; SM00060; FN3; 14.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS00023; FIBRONECTIN_2; 2.
DR PROSITE; PS01253; FIBRONECTIN_1; 12.
KW Glycoprotein; Plasma; Heparin-binding; Acute phase; Phosphorylation;
KW Sulfation; Cell adhesion; Repeat; Alternative splicing;
KW Pyrolydione carboxylic acid.
FT MOD_RES 1 1 PYROLYDIONE CARBOXYLIC ACID.
FT DOMAIN 21 241 FIBRIN- AND HEPARIN-BINDING 1.
FT DOMAIN 277 577 COLLAGEN-BINDING.
FT DNA_BIND 876 1141 CELL-ATTACHMENT.
FT DOMAIN 1236 1509 HEPARIN-BINDING 2.
FT DOMAIN 1600 1870 FIBRIN-BINDING 2.
FT DOMAIN 1991 2216 FIBRONECTIN TYPE-I 1.
FT DOMAIN 19 59 FIBRONECTIN TYPE-I 2.
FT DOMAIN 64 107 FIBRONECTIN TYPE-I 3.
FT DOMAIN 108 151 FIBRONECTIN TYPE-I 4.
FT DOMAIN 153 197 FIBRONECTIN TYPE-I 5.
FT DOMAIN 198 242 FIBRONECTIN TYPE-I 6.
FT DOMAIN 275 314 FIBRONECTIN TYPE-II 1.
FT DOMAIN 314 373 FIBRONECTIN TYPE-II 2.
FT DOMAIN 374 438

```

FT DOMAIN 437 480 FIBRONECTIN TYPE-I 7.  
 FT DOMAIN 485 527 FIBRONECTIN TYPE-I 8.  
 FT DOMAIN 528 571 FIBRONECTIN TYPE-I 9.  
 FT DOMAIN 578 669 FIBRONECTIN TYPE-III 1.  
 FT DOMAIN 688 778 FIBRONECTIN TYPE-III 2.  
 FT DOMAIN 779 874 FIBRONECTIN TYPE-III 3.  
 FT DOMAIN 875 964 FIBRONECTIN TYPE-III 4.  
 FT DOMAIN 965 1054 FIBRONECTIN TYPE-III 5.  
 FT DOMAIN 1055 1141 FIBRONECTIN TYPE-III 6.  
 FT DOMAIN 1142 1234 FIBRONECTIN TYPE-III 7.  
 FT DOMAIN 1235 1325 FIBRONECTIN TYPE-III 8.  
 FT DOMAIN 1326 1415 FIBRONECTIN TYPE-III 9.  
 FT DOMAIN 1416 1509 FIBRONECTIN TYPE-III 10.  
 FT DOMAIN 1510 1599 FIBRONECTIN TYPE-III 11.  
 FT DOMAIN 1600 1691 FIBRONECTIN TYPE-III 12.  
 FT DOMAIN 1692 1780 FIBRONECTIN TYPE-III 13.  
 FT DOMAIN 1781 1870 FIBRONECTIN TYPE-III 14.  
 FT DOMAIN 1871 1990 CONNECTING STRAND 3 (CS-3) (V REGION).  
 FT DOMAIN 1982 2061 FIBRONECTIN TYPE-III 15.  
 FT DOMAIN 2083 2127 FIBRONECTIN TYPE-I 10.  
 FT DOMAIN 2128 2170 FIBRONECTIN TYPE-I 11.  
 FT DOMAIN 2172 2215 FIBRONECTIN TYPE-I 12.  
 FT SITE 1493 1495 CELL ATTACHMENT SITE.

FT DISULFID 21 47  
 FT DISULFID 45 56  
 FT DISULFID 66 94  
 FT DISULFID 92 104  
 FT DISULFID 110 138  
 FT DISULFID 136 148  
 FT DISULFID 155 184  
 FT DISULFID 182 194  
 FT DISULFID 200 229  
 FT DISULFID 227 239  
 FT DISULFID 277 304  
 FT DISULFID 302 311  
 FT DISULFID 329 355  
 FT DISULFID 343 370  
 FT DISULFID 389 415  
 FT DISULFID 403 430  
 FT DISULFID 439 467  
 FT DISULFID 465 477  
 FT DISULFID 487 514  
 FT DISULFID 512 524  
 FT DISULFID 530 558  
 FT DISULFID 556 568  
 FT DISULFID 2085 2114  
 FT DISULFID 2112 2124  
 FT DISULFID 2130 2157  
 FT DISULFID 2155 2167  
 FT DISULFID 2174 2200  
 FT DISULFID 2198 2209  
 FT DISULFID 2246 2246  
 FT DISULFID 2250 2250  
 FT MOD\_RES 845 845  
 FT MOD\_RES 850 850  
 FT CARBOHYD 399 399  
 FT CARBOHYD 497 497  
 FT CARBOHYD 511 511  
 FT CARBOHYD 846 846  
 FT CARBOHYD 976 976  
 FT CARBOHYD 1213 1213  
 FT CARBOHYD 1987 1987  
 FT CARBOHYD 1943 1943  
 FT CARBOHYD 1944 1944  
 FT MOD\_RES 2263 2263  
 SQ SEQUENCE 2265 AA; 249557 MW; C2D21D486F498D5C CRC64;

Query Match 79.5%; Score 35; DB 1; Length 2265;  
 Best Local Similarity 75.0%; Pred. No. 6.4e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 1818 KEGSPPRE 1825  
 RESULT 22  
 FINE\_HUMAN STANDARD; PRT; 2386 AA.  
 AC P02751; Q95609; Q95610; Q14312; Q14325; Q14326; Q86T27; Q8TV18;  
 AC Q96KP7; Q96KP8; Q96KP9; Q9H1B8; Q9HAP3; Q9UMK2;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Fibronectin precursor (FN) (Cold-insoluble globulin) (ClG).  
 GN FNI OR FN.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX MEDLINE=21600194; PubMed=11737888;  
 RA Schor S.L., Schor A.M.;  
 RT "Phenotypic and genetic alterations in mammary stroma: implications for tumour progression.";  
 RL Breast Cancer Res. 3:373-379(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 3; 7 AND 10).  
 RC TISSUE=Cervix;  
 RA Ansoorge W., Krieger S., Regiert T., Rittmuller C., Schwager B.,  
 RA Mewes H.-W., Weil B., Amid C., Osanger A., Fobo G., Han M.,  
 RA Wiemann S.;  
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1-38 FROM N.A.  
 RX MEDLINE=87030890; PubMed=3770189;  
 RA Gutman A., Yamada K.M., Kornblith A.R.;  
 RT "Human fibronectin is synthesized as a pre-propolypeptide.";  
 RL FEBS Lett. 207:145-148(1986).  
 RN [4]  
 RP SEQUENCE OF 1-49 FROM N.A.  
 RX MEDLINE=87175578; PubMed=3031656;  
 RA Dean D.C., Bowls C.L., Bourgeois S.;  
 RT "Cloning and analysis of the promoter region of the human fibronectin gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:1876-1880(1987).  
 RN [5]  
 RP SEQUENCE OF 28-2386 FROM N.A. (ISOFORM 3).  
 RX MEDLINE=85284965; PubMed=2992939;  
 RA Kornblith A.R., Umezawa K., Vibe-Pedersen K., Baralle F.E.;  
 RT "Primary structure of human fibronectin: differential splicing may generate at least 10 polypeptides from a single gene.";  
 RL EMBO J. 4:1755-1759(1985).  
 RN [6]  
 RP SEQUENCE OF 103-481 AND 2228-2386 FROM N.A. (ISOFORMS 1; 3; 8 AND 9).  
 RC TISSUE=Peripheral blood T-cell, and umbilical vein endothelial cells;  
 RA Godfrey H.P., Ebrahim A.A.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE OF 973-2386 FROM N.A. (ISOFORM 3).  
 RX MEDLINE=84272258; PubMed=6462919;  
 RA Kornblith A.R., Vibe-Pedersen K., Baralle F.E.;  
 RT "Human fibronectin: cell specific alternative mRNA splicing generates polypeptide chains differing in the number of internal repeats.";  
 RL Nucleic Acids Res. 12:5853-5868(1984).  
 RN [8]  
 RP SEQUENCE OF 1232-1782 FROM N.A. (ISOFORM 7).  
 RX MEDLINE=88233940; PubMed=3375063;  
 RA Paolletta G., Henchcliffe C., Sebastio G., Baralle F.E.;  
 RT "Sequence analysis and in vivo expression show that alternative splicing of ED-B and ED-A regions of the human fibronectin gene are independent events.";  
 RL Nucleic Acids Res. 16:3545-3557(1988).  
 RN [9]  
 RP SEQUENCE OF 1257-1365 FROM N.A. (ISOFORM 11).



RX MEDLINE=8041070; PubMed=3478690;  
 RA Gutman A., Kornblant A.R.;  
 RT "Identification of a third region of cell-specific alternative  
 RT splicing in human fibronectin mRNA."; *Proc. Natl. Acad. Sci. U.S.A.* 84:7179-7182(1987).  
 RL [10]  
 RN  
 RX MEDLINE=82265604; PubMed=7050098;  
 RA Pierzbacher M.D., Ruoslahti E., Sundelin J., Lind P., Peterson P.A.;  
 RT "The cell attachment domain of fibronectin. Determination of the  
 RT primary structure."; *J. Biol. Chem.* 257:9593-9597(1982).  
 RL [11]  
 RN  
 RX MEDLINE=83290929; PubMed=6688418;  
 RA Oldberg A., Linney E., Ruoslahti E.;  
 RT "Molecular cloning and nucleotide sequence of a cDNA clone coding for  
 RT the cell attachment domain in human fibronectin."; *J. Biol. Chem.* 258:10193-10196(1983).  
 RL [12]  
 RN  
 RX MEDLINE=8611901; PubMed=3003095;  
 RA Oldberg A., Ruoslahti E.;  
 RT "Evolution of the fibronectin gene. Exon structure of cell attachment  
 RT domain."; *J. Biol. Chem.* 261:2113-2116(1986).  
 RL [13]  
 RN  
 RX MEDLINE=85280409; PubMed=292573;  
 RA Bernard M.P., Kolbe M., Weil D., Chu M.-L.;  
 RT "Human cellular fibronectin: comparison of the carboxyl-terminal  
 RT portion with rat identifies primary structural domains separated by  
 RT hypervariable regions."; *Biochemistry* 24:2698-2704(1985).  
 RL [14]  
 RN  
 RX MEDLINE=87026578; PubMed=3021206;  
 RA Sekiguchi K., Kloe A.M., Kurachi K., Yoshitake S., Hakomori S.;  
 RT "Human liver fibronectin complementary DNAs: identification of two  
 RT different messenger RNAs possibly encoding the alpha and beta  
 RT subunits of plasma fibronectin."; *Biochemistry* 25:4936-4941(1986).  
 RL [15]  
 RN  
 RX MEDLINE=22126816; PubMed=12127832;  
 RA Parker A.E., Boutell J., Carr A., Maciewicz R.A.;  
 RT "Novel cartilage-specific splice variants of fibronectin.";  
 RT *Osteoarthritis Cartilage* 10:528-534(2002).  
 RL [16]  
 RN  
 RX MEDLINE=84032463; PubMed=6630202;  
 RA Garcia-Parado A., Pearlstein B., Frangione B.;  
 RT "Primary structure of human plasma fibronectin. The 29,000-dalton  
 RT NH2-terminal domain."; *J. Biol. Chem.* 258:12670-12674(1983).  
 RL [17]  
 RN  
 RX MEDLINE=87080265; PubMed=3024962;  
 RA Owens R.J., Baralle F.E.;  
 RT "Mapping the collagen-binding site of human fibronectin by expression  
 RT in *Escherichia coli*."; *EMBO J.* 5:2825-2830(1986).  
 RL [18]  
 RN  
 RX MEDLINE=86042625; PubMed=2414772;  
 RA Liu M.C., Yu S., Sy J., Redman C.M., Lipmann F.;  
 RT "Tyrosine sulfation of proteins from the human hepatoma cell line  
 RT HepG2."; *Proc. Natl. Acad. Sci. U.S.A.* 82:7160-7164(1985).  
 RL [19]  
 RN  
 RX MEDLINE=91190085; PubMed=2012601;  
 RT O-GLYCOSYLATION OF THR-2064.

RA Tresselt T., McCarthy J.B., Calaycay J., Lee T.D., Legesse K.,  
 RA Shively J.E., Pande H.;  
 RT "Human plasma fibronectin. Demonstration of structural differences  
 RT between the A- and B-chains in the III CS region."; *Biochem. J.* 274:731-738(1991).  
 RL [20]  
 RN  
 RX MEDLINE=93015879; PubMed=1400330;  
 RA Balbona K., Tran H., Godyna S., Ingham K.C., Strickland D.K.,  
 RA Agraves W.S.;  
 RT "Fibulin binds to itself and to the carboxyl-terminal heparin-binding  
 RT region of fibronectin."; *J. Biol. Chem.* 267:20120-20125(1992).  
 RL [21]  
 RN  
 RX MEDLINE=95081153; PubMed=7989369;  
 RA Rostagno A., Williams M.J., Baron M., Campbell I.D., Gold L.I.;  
 RT "Further characterization of the NH2-terminal fibrin-binding site on  
 RT fibronectin."; *J. Biol. Chem.* 269:31938-31945(1994).  
 RL [22]  
 RN  
 RX MEDLINE=92162710; PubMed=1311202;  
 RA Baron M., Main A.L., Driscoll P.C., Mardon H.J., Boyd J.,  
 RA Campbell I.D.;  
 RT "1H NMR assignment and secondary structure of the cell adhesion type  
 RT III module of fibronectin."; *Biochemistry* 31:2068-2073(1992).  
 RL [23]  
 RN  
 RX MEDLINE=9304665; PubMed=1423622;  
 RA Main A.L., Harvey T.S., Baron M., Boyd J., Campbell I.D.;  
 RT "The three-dimensional structure of the tenth type III module of  
 RT fibronectin: an insight into RGD-mediated interactions."; *Cell* 71:671-678(1992).  
 RL [24]  
 RN  
 RX MEDLINE=94141923; PubMed=8308892;  
 RA Williams M.J., Phan I., Harvey T.S., Rostagno A., Gold L.I.,  
 RA Campbell I.D.;  
 RT "Solution structure of a pair of fibronectin type I modules with  
 RT fibrin binding activity."; *J. Mol. Biol.* 235:1302-1311(1994).  
 RL [25]  
 RN  
 RX MEDLINE=96069779; PubMed=7583666;  
 RA Potts J.R., Phan I., Williams M.J., Campbell I.D.;  
 RT "High-resolution structural studies of the factor XIIIa crosslinking  
 RT site and the first type I module of fibronectin."; *Nat. Struct. Biol.* 2:946-950(1995).  
 RL [26]  
 RN  
 RX MEDLINE=98179558; PubMed=9514732;  
 RA Sticht H., Pickford A.R., Potts J.R., Campbell I.D.;  
 RT "Solution structure of the glycosylated second type 2 module of  
 RT fibronectin."; *J. Mol. Biol.* 276:177-187(1998).  
 RL [27]  
 RN  
 RX STRUCTURE BY NMR OF EXTRA ED-B DOMAIN FROM ISOFORM 7.

Query Match 79.5%; Score 35; DB 1; Length 2386;  
 Best Local Similarity 75.0%; Pred. No. 6.7e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KPSPPEE 8  
 Db 1939 KPSPPEE 1946

RESULT 23  
 ID P1NC\_MOUSE STANDARD; PRT; 2477 AA.  
 AC P11276; Q61567; Q61568; Q61569; Q64233; Q80U14;

DT 01-JUL-1989 (Rel. 11, Created)  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Fibronectin precursor (FN).  
 GN FN1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE OF 1-920 FROM N.A.  
 RC STRAIN=FVB/N-3; TISSUE=Breast tumor;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Klausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heist F.,  
 RA Diatchenko L., Marinska K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE OF 1-28 FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=94131313; PubMed=8299972;  
 RA Polly P., Nicholson R.C.;  
 RT "Sequence of the mouse fibronectin-encoding gene promoter region."  
 RL Gene 137:353-354(1993).  
 RN [3]  
 RP SEQUENCE OF 562-834 FROM N.A.  
 RC STRAIN=MNR1;  
 RX MEDLINE=95403556; PubMed=7673336;  
 RA Tait J.F., Weller A., Timpl R., Ekblom M., Ekblom P.;  
 RT "Regulation of mesenchymal extracellular matrix protein synthesis by  
 RT transforming growth factor-beta and glucocorticoids in tumor  
 RT stroma."  
 RL J. Cell Sci. 108:2153-2162(1995).  
 RN [4]  
 RP SEQUENCE OF 899-2376 FROM N.A.  
 RC Gorski G., Aros M., Norton P.;  
 RX Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 RL [5]  
 RP SEQUENCE OF 2375-2477 FROM N.A.  
 RX MEDLINE=8124987; PubMed=3124113;  
 RA Blatti S.P., Foster D.N., Ranganathan G., Moses H.L., Getz M.J.;  
 RT "Induction of fibronectin gene transcription and mRNA is a primary  
 RT response to growth-factor stimulation of AKR-2B cells."  
 RL Proc. Natl. Acad. Sci. U.S.A. 85:1119-1123(1988).  
 RN [6]  
 RP SEQUENCE OF 2375-2477 FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=93011702; PubMed=1327855;  
 RA Khandjian E.W., Salomon C., Leonard N., Tremblay S., Turlet H.;  
 RT "Fibronectin gene expression in proliferating, quiescent, and SV40-  
 RT infected mouse kidney cells."  
 RL Exp. Cell Res. 202:464-470(1992).  
 RN [7]  
 RP STRUCTURE BY NMR OF 1447-1630.  
 RX MEDLINE=98202578; PubMed=9533887;  
 RA Copie V., Tomita Y., Akiyama S.K., Aota S., Yamada K.M., Venable R.M.,  
 RA Pastor R.W., Krueger S., Torchia D.A.;

RT "solution structure and dynamics of linked cell attachment modules of  
 RT mouse fibronectin containing the RGD and synergy regions: comparison  
 RT with the human fibronectin crystal structure."  
 RL J. Mol. Biol. 277:663-682(1998).  
 RN [8]  
 RP DOWN-REGULATION BY GLUCOCORTICOID.  
 RX MEDLINE=21600963; PubMed=11737251;  
 RA Gu Y.-C., Tait J.F., Gullberg D., Timpl R., Ekblom M.;  
 RT "Glucocorticoids down-regulate the extracellular matrix proteins  
 RT fibronectin, fibulin-1 and fibulin-2 in bone marrow stroma."  
 RL Eur. J. Haematol. 67:176-184(2001).  
 CC -1- FUNCTION: Fibronectins bind cell surfaces and various compounds  
 CC including collagen, fibrin, heparin, DNA, and actin. Fibronectins  
 CC are involved in cell adhesion, cell motility, opsonization, wound  
 CC healing, and maintenance of cell shape.  
 CC -1- SUBUNIT: Mostly heterodimers or multimers of alternatively spliced  
 CC variants, connected by 2 disulfide bonds near the carboxyl ends;  
 CC to a lesser extend homodimers. Interacts with PDN1 (By  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Comment=A number of isoforms are produced. Each of the "extra  
 CC domain" and the connecting strand 3 are present in some forms of  
 CC fibronectin and absent in others;  
 CC Name=1;  
 CC IsoId=P11276-1; Sequence=Displayed;  
 CC -1- TISSUE SPECIFICITY: Plasma FN (soluble dimeric form) is secreted  
 CC by hepatocytes. Cellular FN (dimeric or cross-linked multimeric  
 CC forms), made by fibroblasts, epithelial and other cell types, is  
 CC deposited as fibrils in the extracellular matrix.  
 CC -1- INDUCTION: Glucocorticoids suppressed mRNA expression and protein  
 CC synthesis.  
 CC -1- PTM: Sulfated (By similarity).  
 CC -1- SIMILARITY: Contains 12 fibronectin type I domains.  
 CC -1- SIMILARITY: Contains 12 fibronectin type II domains.  
 CC -1- SIMILARITY: Contains 17 fibronectin type III domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: BC051082; AAH51082.1; -.  
 DR EMBL: Z22729; CAH80422.1; -.  
 DR EMBL: X82402; CAA57796.1; -.  
 DR EMBL: X93167; CAA63654.1; -.  
 DR EMBL: M8194; AAA37636.1; -.  
 DR EMBL: S45680; AAB23491.1; -.  
 DR PIR: A49173; AA9173.  
 DR PIR: I48349; I48349.  
 DR PDB: 1MFN; 29-APR-98.  
 DR PDB: 2MFN; 29-APR-98.  
 DR MGI: MGI:95566; Fn1.  
 DR GO: GO:0007155; P:cell adhesion; IDA.  
 DR InterPro: IPR006209; EGF like.  
 DR InterPro: IPR000083; Fibrinchn.  
 DR InterPro: IPR008957; FN\_III-like.  
 DR InterPro: IPR003961; FN\_III.  
 DR InterPro: IPR000562; FN\_Type\_II.  
 DR InterPro: IPR003962; Fn1I subd.  
 DR InterPro: IPR008924; MCR\_alpha\_beta\_C.  
 DR PRINTS: PRO0014; EWTYPER11.  
 DR SMART: SM00058; FN1; 4.  
 DR SMART: SM00060; FN3; 12.  
 DR PROSITE: PS00022; EGF\_1; 2.  
 DR PROSITE: PS00023; FIBRONECTIN\_2; 2.  
 DR PROSITE: PS01253; FIBRONECTIN\_1; 12.  
 KW Glycoprotein; Plasma; Heparin-binding; Acute phase; Phosphorylation;  
 KW Sulfation; Cell adhesion; Repeat; Alternative splicing; Signal;

3d-structure. 1 32  
 FT SIGNAL 1 32  
 FT CHAIN 1 32  
 FT DOMAIN 33 2477  
 FT DOMAIN 53 273  
 FT DOMAIN 308 608  
 FT DNA\_BIND 906 1171  
 FT DOMAIN 1357 1630  
 FT DOMAIN 1811 2081  
 FT DOMAIN 2296 2427  
 FT DOMAIN 51 36  
 FT DOMAIN 96 140  
 FT DOMAIN 140 185  
 FT DOMAIN 185 230  
 FT DOMAIN 230 272  
 FT DOMAIN 306 343  
 FT DOMAIN 345 404  
 FT DOMAIN 405 469  
 FT DOMAIN 468 516  
 FT DOMAIN 516 559  
 FT DOMAIN 559 602  
 FT DOMAIN 609 706  
 FT DOMAIN 707 808  
 FT DOMAIN 809 903  
 FT DOMAIN 904 994  
 FT DOMAIN 995 1084  
 FT DOMAIN 1085 1172  
 FT DOMAIN 1173 1264  
 FT DOMAIN 1265 1355  
 FT DOMAIN 1356 1446  
 FT DOMAIN 1447 1536  
 FT DOMAIN 1537 1630  
 FT DOMAIN 1631 1720  
 FT DOMAIN 1721 1810  
 FT DOMAIN 1811 1902  
 FT DOMAIN 1903 1991  
 FT DOMAIN 1992 2081  
 FT DOMAIN 2082 2201  
 FT DOMAIN 2202 2283  
 FT DOMAIN 2294 2338  
 FT DOMAIN 2339 2381  
 FT DOMAIN 2383 2426  
 FT SITE 1614 1616  
 FT SITE 2181 2183  
 FT DISULFID 53 79  
 FT DISULFID 77 88  
 FT DISULFID 98 126  
 FT DISULFID 124 136  
 FT DISULFID 142 170  
 FT DISULFID 168 180  
 FT DISULFID 187 216  
 FT DISULFID 214 226  
 FT DISULFID 232 261

Query Match 79.5%; Score 35; DB 1; Length 2477;  
 Best Local Similarity 75.0%; Pred. No. 7e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
 DB 2029 KPSSPPE 2036

RESULT 24  
 PINC\_RAT STANDARD; PRT; 2477 AA.  
 AC P04937;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Fibronectin precursor (FN).  
 GN FNI.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_Taxid=10116;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Fischer; TISSUE=Liver;  
 RX MEDLINE=8054951; PubMed=2445560;  
 RA Schwarzbauer J.E., Patel R.S., Ponda D., Hynes R.O.;  
 RT "Multiple sites of alternative splicing of the rat fibronectin gene transcript.";  
 RL EMO J. 6:2573-2580(1987).  
 RN [2]  
 RP SEQUENCE OF 1-139 AND 2382-2477 FROM N.A.  
 RC STRAIN=Fischer; TISSUE=Liver;  
 RX MEDLINE=8054950; PubMed=3119323;  
 RA Patel R.S., Odermatt E., Schwarzbauer J.E., Hynes R.O.;  
 RT "Organization of the fibronectin gene provides evidence for exon shuffling during evolution.";  
 RL EMO J. 6:2565-2572(1987).  
 RN [3]  
 RP SEQUENCE OF 1586-2477 FROM N.A.  
 RX MEDLINE=84082067; PubMed=6337187;  
 RA Schwarzbauer J.E., Tamkun J.W., Lemischka I.R., Hynes R.O.;  
 RT "Three different fibronectin mRNAs arise by alternative splicing within the coding region.";  
 RL Cell 35:421-431(1983).  
 CC -1- FUNCTION: Fibronectins bind cell surfaces and various compounds including collagen, fibrin, heparin, DNA and actin. Fibronectins are involved in cell adhesion, cell motility, opsonization, wound healing, and maintenance of cell shape.  
 CC -1- SUBUNIT: MOSTLY HETERODIMERS OR MULTIMERS OF ALTERNATIVELY SPLICED VARIANTS, CONNECTED BY 2 DISULFIDE BONDS NEAR THE CARBOXYL ENDS; TO A LESSER EXTEND HOMODIMERS.  
 CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=4;  
 CC Comment=Each of the "extra domain" and the connecting strand 3 are present in some forms of fibronectin and absent in others;  
 CC Name=1;  
 CC IsoId=P04937-1; Sequence=Displayed;  
 CC Name=2; Synonyms=FNIII-13-1ess;  
 CC IsoId=P04937-2; Sequence=VSP\_003258;  
 CC Name=3; Synonyms=Lambda-RLR4-5;  
 CC IsoId=P04937-3; Sequence=VSP\_003259;  
 CC Name=4; Synonyms=Lambda-RLR6;  
 CC IsoId=P04937-4; Sequence=VSP\_003260;  
 CC -1- TISSUE SPECIFICITY: Plasma FN (soluble dimeric form) is secreted by hepatocytes. Cellular FN (dimeric or cross-linked multimeric forms), made by fibroblasts, epithelial and other cell types, is deposited as fibrils in the extracellular matrix.  
 CC -1- PTM: Sulfated (By similarity).  
 CC -1- SIMILARITY: Contains 12 fibronectin type I domains.  
 CC -1- SIMILARITY: Contains 2 fibronectin type II domains.  
 CC -1- SIMILARITY: Contains 17 fibronectin type III domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X15906; CAA34020.1; -;  
 DR EMBL; L29191; AAA41166.1; -;  
 DR EMBL; L29191; AAA41166.1; JOINED.  
 DR EMBL; L29191; AAA41167.1; -;  
 DR EMBL; L29191; AAA41167.1; JOINED.  
 DR EMBL; L29191; AAA41168.1; -;  
 DR EMBL; L00191; AAA41168.1; JOINED.  
 DR EMBL; X05831; CAA29278.1; -;  
 DR EMBL; X05831; CAA29279.1; -;  
 DR EMBL; X05833; CAA29280.1; -;  
 DR EMBL; X05834; CAA29281.1; -;

DR	PIR, S14428; S14428.	
DR	HSSP; P02751; 1FBR.	
DR	InterPro; IPR006209; EGF_like.	
DR	InterPro; IPR008957; FN_III-like.	
DR	InterPro; IPR008957; FN_III-like.	
DR	InterPro; IPR003961; FN_III.	
DR	InterPro; IPR005652; FN_Type_II.	
DR	InterPro; IPR003962; FnIII_subd.	
DR	Pfam; PF000039; fn1; 12.	
DR	Pfam; PF000040; fn2; 2.	
DR	Pfam; PF00041; fn3; 17.	
DR	PRINTS; PR00013; ENTPEB11.	
DR	PRINTS; PR00014; ENTPEB11.	
DR	ProDom; PD0000995; FN_Type_II; 2.	
DR	SMART; SM00058; FN1; 12.	
DR	SMART; SM00059; FN2; 2.	
DR	SMART; SM00060; FN3; 13.	
DR	PROSITE; PS00022; EGF_1; 2.	
DR	PROSITE; PS00023; FIBRONECTIN_2; 2.	
DR	PROSITE; PS01233; FIBRONECTIN_1; 12.	
KM	GlycoProfile; Plasma; Heparin-binding; Acute phase; Phosphorylation; Sulfation; Cell adhesion; Repeat; Alternative splicing; Signal.	
KM	SIGNAL; Cell adhesion; Repeat; Alternative splicing; Signal.	
FT	CHAIN	1 32
FT	DOMAIN	33 2477
FT	DOMAIN	53 273
FT	DNA_BIND	308 608
FT	DOMAIN	906 1171
FT	DOMAIN	1357 1630
FT	DOMAIN	1811 2081
FT	DOMAIN	2296 2427
FT	DOMAIN	51 91
FT	DOMAIN	96 139
FT	DOMAIN	140 183
FT	DOMAIN	185 229
FT	DOMAIN	230 274
FT	DOMAIN	306 345
FT	DOMAIN	345 404
FT	DOMAIN	405 469
FT	DOMAIN	468 511
FT	DOMAIN	516 558
FT	DOMAIN	559 602
FT	DOMAIN	609 706
FT	DOMAIN	707 808
FT	DOMAIN	809 903
FT	DOMAIN	904 994
FT	DOMAIN	995 1084
FT	DOMAIN	1085 1172
FT	DOMAIN	1173 1264
FT	DOMAIN	1265 1355
FT	DOMAIN	1356 1446
FT	DOMAIN	1447 1536
FT	DOMAIN	1537 1630
FT	DOMAIN	1631 1720
FT	DOMAIN	1721 1810
FT	DOMAIN	1811 1902
FT	DOMAIN	1903 1991
FT	DOMAIN	1992 2081
FT	DOMAIN	2082 2201
FT	DOMAIN	2202 2283
FT	DOMAIN	2294 2338
FT	DOMAIN	2339 2381
FT	DOMAIN	2383 2426
FT	DOMAIN	2427 2518
FT	DOMAIN	2519 2610
FT	DOMAIN	2611 2702
FT	DOMAIN	2703 2794
FT	DOMAIN	2795 2886
FT	DOMAIN	2887 2978
FT	DOMAIN	2979 3070
FT	DOMAIN	3071 3162
FT	DOMAIN	3163 3254
FT	DOMAIN	3255 3346
FT	DOMAIN	3347 3438
FT	DOMAIN	3439 3530
FT	DOMAIN	3531 3622
FT	DOMAIN	3623 3714
FT	DOMAIN	3715 3806
FT	DOMAIN	3807 3898
FT	DOMAIN	3899 3990
FT	DOMAIN	3991 4082
FT	DOMAIN	4083 4174
FT	DOMAIN	4175 4266
FT	DOMAIN	4267 4358
FT	DOMAIN	4359 4450
FT	DOMAIN	4451 4542
FT	DOMAIN	4543 4634
FT	DOMAIN	4635 4726
FT	DOMAIN	4727 4818
FT	DOMAIN	4819 4910
FT	DOMAIN	4911 5002
FT	DOMAIN	5003 5094
FT	DOMAIN	5095 5186
FT	DOMAIN	5187 5278
FT	DOMAIN	5279 5370
FT	DOMAIN	5371 5462
FT	DOMAIN	5463 5554
FT	DOMAIN	5555 5646
FT	DOMAIN	5647 5738
FT	DOMAIN	5739 5830
FT	DOMAIN	5831 5922
FT	DOMAIN	5923 6014
FT	DOMAIN	6015 6106
FT	DOMAIN	6107 6198
FT	DOMAIN	6199 6290
FT	DOMAIN	6291 6382
FT	DOMAIN	6383 6474
FT	DOMAIN	6475 6566
FT	DOMAIN	6567 6658
FT	DOMAIN	6659 6750
FT	DOMAIN	6751 6842
FT	DOMAIN	6843 6934
FT	DOMAIN	6935 7026
FT	DOMAIN	7027 7118
FT	DOMAIN	7119 7210
FT	DOMAIN	7211 7302
FT	DOMAIN	7303 7394
FT	DOMAIN	7395 7486
FT	DOMAIN	7487 7578
FT	DOMAIN	7579 7670
FT	DOMAIN	7671 7762
FT	DOMAIN	7763 7854
FT	DOMAIN	7855 7946
FT	DOMAIN	7947 8038
FT	DOMAIN	8039 8130
FT	DOMAIN	8131 8222
FT	DOMAIN	8223 8314
FT	DOMAIN	8315 8406
FT	DOMAIN	8407 8498
FT	DOMAIN	8499 8590
FT		

FT	DISULFID	259	271		BY SIMILARITY.
FT	DISULFID	308	335		BY SIMILARITY.
FT	DISULFID	333	342		BY SIMILARITY.
FT	DISULFID	360	386		BY SIMILARITY.
FT	DISULFID	374	401		BY SIMILARITY.
FT	DISULFID	420	446		BY SIMILARITY.
FT	DISULFID	434	461		BY SIMILARITY.
FT	DISULFID	470	498		BY SIMILARITY.
FT	DISULFID	496	508		BY SIMILARITY.
FT	DISULFID	518	545		BY SIMILARITY.
FT	DISULFID	543	555		BY SIMILARITY.
FT	DISULFID	561	589		BY SIMILARITY.
FT	DISULFID	587	599		BY SIMILARITY.
FT	DISULFID	2296	2325		BY SIMILARITY.
FT	DISULFID	2323	2335		BY SIMILARITY.
FT	DISULFID	2341	2368		BY SIMILARITY.
FT	DISULFID	2366	2378		BY SIMILARITY.
FT	DISULFID	2385	2409		BY SIMILARITY.
FT	DISULFID	2407	2423		BY SIMILARITY.
FT	DISULFID	2458	2458		INTERCHAIN (WITH C-2462).
FT	DISULFID	2462	2462		INTERCHAIN (WITH C-2458).
FT	MOD_RES	875	875		SULFATION (POTENTIAL).
FT	MOD_RES	880	880		SULFATION (POTENTIAL).
FT	MOD_RES	2392	2392		SULFATION (POTENTIAL).
FT	MOD_RES	2475	2475		SULFATION (POTENTIAL).
FT	CARBOHYD	430	430		PHOSPHORYLATION (BY SIMILARITY).
FT	CARBOHYD	528	528		N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT	CARBOHYD	542	542		N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT	CARBOHYD	876	876		N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT	CARBOHYD	1006	1006		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1243	1243		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2198	2198		N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT	CARBOHYD	2154	2154		O-LINKED (GALNAC. . .) (BY SIMILARITY).
FT	VARSPLIC	1720	1809		Missing (in isoform 2).
FT	VARSPLIC	2082	2106		/FtId=VSP_003258.
FT	VARSPLIC	2082	2106		Missing (in isoform 3).
FT	VARSPLIC	2082	2200		/FtId=VSP_003259.
FT	CONFLICT	2318	2318		Missing (in isoform 4).
FT	SEQUENCE	2477 AA;	272510 MW;	B491A1472CEBDEB5 CRC64;	G -> A (IN REF. 3).
QY	1 KPSSPPEE	8			
DB	2029 KPGSPPRE	2036			
RESULT 25					
ID	GBF3_ARATH	STANDARD;	PRT;	362 AA.	
AC	P42776;				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	15-DEC-1998 (Rel. 37, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	G-box binding factor 3.				
GN	GBF3 OR AT2G46270 OR TGF17.8.				
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;				
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.				
OX	NCBI_TaxId=3702;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=cv. Columbia;				
RX	MEDLINE=96240020; PubMed=8672884;				
RT	Iu G.H., Paul A.L., McCarty D.R., Perl R.J.;				
TL	"Transcription factor vcracity: is GBF3 responsible for ABA-regulated expression of Arabidopsis Adh?";				
XL	Plant Cell 8:847-857(1996).				

RN [2] SEQUENCE FROM N.A.  
 RP STRAIN=cv. Columbia;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Macon T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Roming C.M., Koo H.L.,  
 RA Moffett K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umeyan L.,  
 RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creaey T.H.,  
 RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,  
 RA Niehan W.C., White O., Bisen J.A., Salzberg S.L., Frazer C.M.,  
 RA Venter J.C.;  
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 402:761-768(1999).  
 RN [3] SEQUENCE OF 59-382 FROM N.A.  
 RP STRAIN=cv. Columbia; TISSUE=leaf, and stem;  
 RX MEDLINE=92224864; PubMed=1373374;  
 RA Schindler U., Menkens A.E., Beckmann H., Ecker J.R., Cashmore A.R.;  
 RT "Heterodimerization between light-regulated and ubiquitously  
 RT expressed Arabidopsis GBR bZIP proteins.";  
 RL EMO J. 11:1261-1273(1992).  
 CC -1- FUNCTION: Binds to the G-box motif (5'-CCACGTG-3') of the rbcS-1A  
 CC gene promoter. G-box and G-box-like motifs are cis-acting elements  
 CC defined in promoters of certain plant genes which are regulated by  
 CC such diverse stimuli as light-induction or hormone control.  
 CC -1- SUBUNIT: DNA-binding heterodimer.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: Present only in dark grown leaves and roots.  
 CC -1- SIMILARITY: Belongs to the bZIP family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, U17891; AAA0947.1; -  
 CC EMBL, U17850; AAB0611.1; -  
 CC EMBL, AC005397; AAC62879.1; -  
 CC EMBL, X63896; CAA5358.1; -  
 CC PIR, G84900; G84900.  
 CC PIR, S20885; S20885.  
 CC HSSP, P03069; IYSA.  
 CC TRANSFAC, T01080; -  
 CC InterPro, IPR004827; TF\_bZIP.  
 CC Pfam, PF00170; bZIP.1.  
 CC SMART, SM00336; BRLZ.1.  
 CC PROSITE, PS50217; bZIP.1.  
 CC PROSITE, PS00036; bZIP\_BASIC.1.  
 CC Transcription regulation; DNA-binding; Nuclear protein.  
 CC FT DOMAIN 8 121 PRO-RICH.  
 CC FT DNA BIND 261 280 BASIC MOTIF.  
 CC FT DOMAIN 287 315 LECICINE-ZIPPER.  
 CC FT CONFLICT 75 75 Y->S (IN REF. 1).  
 CC FT CONFLICT 366 366 L->P (IN REF. 1).  
 CC SQ SEQUENCE 382 AA; 41114 MW; 48B3176ZCDD977C CRC64;  
 Query Match 77.3%; Score 34; DB 1; Length 382;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

AC P34960;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Macrophage metalloelastase precursor (EC 3.4.24.65) (MME) (Matrix  
 DE metalloproteinase-12) (MMP-12).  
 GN MMP12 OR MME1 OR MME.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1] SEQUENCE FROM N.A., AND SEQUENCE OF 99-125.  
 RP TISSUE=Macrophage;  
 RX MEDLINE=92165826; PubMed=1537850;  
 RA Shapiro S.D., Griffin G.L., Gilbert D.J., Jenkins N.A.,  
 RA Copeland N.G., Welgus H.G., Senior R.M., Ley T.T.;  
 RT "Molecular cloning, chromosomal localization, and bacterial  
 RT expression of a murine macrophage metalloelastase.";  
 RL J. Biol. Chem. 267:4664-4671(1992).  
 CC -1- FUNCTION: May be involved in tissue injury and remodeling. Has  
 CC significant elastolytic activity.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of soluble and insoluble elastin.  
 CC Specific cleavages are also produced at 14-Ala|-Leu-15 and 16-  
 CC Tyr|-Leu-17 in the B chain of insulin.  
 CC -1- COFACTOR: Binds 2 zinc ions and 4 calcium ions per subunit (By  
 CC similarity).  
 CC -1- SIMILARITY: Belongs to peptidase family M10A.  
 CC -1- SIMILARITY: Contains 1 hemopexin-like domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, M82831; AAA39526.1; -  
 CC PIR, A42401; A42401.  
 CC HSSP, P03956; 1CGL.  
 CC MGD, MGI:97005; Mmp12.  
 CC InterPro, IPR000585; Hemopexin.  
 CC InterPro, IPR001818; Pept\_M10A\_M12B.  
 CC InterPro, IPR006025; Pept\_M\_Zn\_BS.  
 CC InterPro, IPR006026; Peptidase\_M.  
 CC Pfam, PF00045; hemopexin.4.  
 CC Pfam, PF00413; Peptidase\_M10; 1.  
 CC Pfam, PF03933; Peptidase\_M10\_N; 1.  
 CC PRINTS, PR00138; MATRXIN.  
 CC SMART, SM00120; HX; 4.  
 CC SMART, SM00235; ZnMC; 1.  
 CC PROSITE, PS00024; HEMOPEXIN.1.  
 CC PROSITE, PS00142; ZINC\_PROTEASE.1.  
 CC PROSITE, PS00546; CYSTEINE\_SWITCH.1.  
 CC Hydrolyase; Metalloprotease; Glycoprotein; Calcium-binding;  
 CC Metal-binding; Zinc; Zymogen; Extracellular matrix; Signal.  
 CC FT SIGNAL 1 17 PROBABLE.  
 CC FT PROPEP 18 98 ACTIVATION PEPTIDE.  
 CC FT CHAIN 99 462 MACROPHAGE METALLOELASTASE.  
 CC FT DOMAIN 272 462 CYSTEINE\_SWITCH (BY SIMILARITY).  
 CC FT SITE 85 85 CALCIUM 1 (BY SIMILARITY).  
 CC FT METAL 117 117 CALCIUM 2 (BY SIMILARITY).  
 CC FT METAL 151 151 ZINC 1 (BY SIMILARITY).  
 CC FT METAL 161 161 ZINC 1 (BY SIMILARITY).  
 CC FT METAL 163 163 CALCIUM 3 (BY SIMILARITY).  
 CC FT METAL 168 168 CALCIUM 3 (BY SIMILARITY).  
 CC FT METAL 169 169 CALCIUM 3 (VIA CARBOXYL OXYGEN) (BY  
 CC SIMILARITY).  
 CC FT METAL 171 171 CALCIUM 3 (VIA CARBOXYL OXYGEN) (BY  
 CC SIMILARITY).  
 CC FT METAL 173 173 CALCIUM 3 (VIA CARBOXYL OXYGEN) (BY

```

FT METAL 176 176 SIMILARITY).
FT METAL 183 183 ZINC 1 (BY SIMILARITY).
FT METAL 187 187 CALCULUM 2 (VIA CARBONYL OXYGEN) (BY
FT METAL 189 189 SIMILARITY).
FT METAL 191 191 ZINC 1 (BY SIMILARITY).
FT METAL 192 192 CALCULUM 3 (BY SIMILARITY).
FT METAL 194 194 CALCULUM 1 (BY SIMILARITY).
FT METAL 211 211 CALCULUM 1 AND 3 (BY SIMILARITY).
FT ACT SITE 212 212 ZINC 2 (CATALYTIC) (BY SIMILARITY).
FT METAL 215 215 ZINC 2 (CATALYTIC) (BY SIMILARITY).
FT METAL 221 221 ZINC 2 (CATALYTIC) (BY SIMILARITY).
FT METAL 282 282 ZINC 2 (CATALYTIC) (BY SIMILARITY).
FT METAL 374 374 ZINC 2 (CATALYTIC) (BY SIMILARITY).
FT METAL 423 423 ZINC 2 (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD 21 21 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 275 462 BY SIMILARITY.
SQ SEQUENCE 462 AA; 53841 MM; BB9625906FIDBDF CRC64;

Query Match 77.3%; Score 34; DB 1; Length 462;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPP 6
DB 266 KPSSPP 271

RESULT 27
GAG_HV1C4 STANDARD; PRT; 499 AA.
ID_GAG_HV1C4
AC P05887;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE GAG polypeptide [contains: Core proteins p17, p24, p7, p1, p6].
GN GAG.
OS Human immunodeficiency virus type 1 (CDC-451 isolate) (HIV-1).
OC Viruses; Retrovirdae; Lentiviridae; Lentivirus.
OX NCBI_Taxid=11687;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=67041461; Pubmed=3490666;
RA Desai S.M., Kalyanaraman V.S., Casey J.M., Srinivasan A.,
RA Andersen P.R., Devare S.G.;
RT "Molecular cloning and primary nucleotide sequence analysis of a
RT distinct human immunodeficiency virus isolate reveal significant
RT divergence in its genomic sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 83:8380-8384(1986).
CC -1- FUNCTION: PERFORMS HIGHLY COMPLEX ORCHESTRATED TASKS DURING THE
CC ASSEMBLY, BUDDING, MATURATION, AND INFECTION STAGES OF THE VIRAL
CC REPLICATION CYCLE. DURING VIRAL ASSEMBLY, THE PROTEINS FORM
CC MEMBRANE ASSOCIATIONS AND SELF-ASSOCIATIONS THAT ULTIMATELY
CC RESULT IN BUDDING OF AN IMMATURE VIRION FROM THE INFECTED CELL.
CC GAG PRECURSORS ALSO FUNCTION DURING VIRAL ASSEMBLY TO SELECTIVELY
CC BIND AND PACKAGE TWO PLUS STRANDS OF GENOMIC RNA.
CC -1- FUNCTION: The p24 protein is phosphorylated.
CC -1- SIMILARITY: Contains 2 CCHC-type zinc fingers.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).

```

```

DR EMBL: M13136; AAA44306.1; -.
DR PIR: A25523; F0VWH4.
DR HSSP: P05888; 1AAF.
DR HIV: M13136; GAGSCD45.
DR InterPro: IPR000721; Gag_p24.
DR InterPro: IPR008916; Retrov_capsid_C.
DR InterPro: IPR008919; Retrov_capsid_N.
DR InterPro: IPR000071; Retrov_p17.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00540; Gag_p17; 1.
DR Pfam: PF00607; Gag_p24; 1.
DR Pfam: PF00098; Zf_CCHC; 2.
DR PRINTS: PR00939; C2HCZNFINGER.
DR PRINTS: PR00234; HIVMATRIX.
DR SMART: SM00343; ZNF_C2HC; 2.
DR PROSITE: PS50158; ZF_CCHC; 2.
KW AIDS: Core protein; Polypeptide; Myristate; Phosphorylation;
KW Zinc-finger; Repeat; Lipoprotein.
FT INIT MET 0
FT CHAIN 1 131 CORE PROTEIN P17 (MATRIX PROTEIN).
FT CHAIN 132 362 CORE PROTEIN P24 (CORE ANTIGEN).
FT CHAIN 363 376 CORE PROTEIN P2.
FT CHAIN 377 431 CORE PROTEIN P7 (NUCLEOCAPSID PROTEIN).
FT CHAIN 432 447 CORE PROTEIN P1.
FT CHAIN 448 499 CORE PROTEIN P6.
FT ZN FING 389 406 CCHC-TYPE 1.
FT ZN FING 410 427 CCHC-TYPE 2.
FT LPID 1 1 N-myristoyl glycine (in host) (By
FT similarity).
SQ SEQUENCE 499 AA; 55796 MM; 023CA76C9C6CF22AD CRC64;

Query Match 77.3%; Score 34; DB 1; Length 499;
Best Local Similarity 62.5%; Pred. No. 1.9e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8
DB 453 EPTAPPE 460

RESULT 28
GAG_HV1H2 STANDARD; PRT; 499 AA.
ID_GAG_HV1H2
AC P04591;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE GAG polypeptide [contains: Core proteins p17, p24, p7, p1, p6].
GN GAG.
OS Human immunodeficiency virus type 1 (HXB2 isolate) (HIV-1).
OC Viruses; Retrovirdae; Lentiviridae; Lentivirus.
OX NCBI_Taxid=11706;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=6729196; Pubmed=3040055;
RA Retner L., Fisher A., Jagodzinski L.L., Mitsuya H., Lion R.-S.,
RA Gallo R.C., Wong-Staal F.;
RT "Complete nucleotide sequences of functional clones of the AIDS
RT virus."
RT AIDS Res. Hum. Retroviruses 3:57-69(1987).
CC -1- FUNCTION: PERFORMS HIGHLY COMPLEX ORCHESTRATED TASKS DURING THE
CC ASSEMBLY, BUDDING, MATURATION, AND INFECTION STAGES OF THE VIRAL
CC REPLICATION CYCLE. DURING VIRAL ASSEMBLY, THE PROTEINS FORM
CC MEMBRANE ASSOCIATIONS AND SELF-ASSOCIATIONS THAT ULTIMATELY
CC RESULT IN BUDDING OF AN IMMATURE VIRION FROM THE INFECTED CELL.
CC GAG PRECURSORS ALSO FUNCTION DURING VIRAL ASSEMBLY TO SELECTIVELY
CC BIND AND PACKAGE TWO PLUS STRANDS OF GENOMIC RNA.
CC -1- FUNCTION: The p24 protein is phosphorylated.
CC -1- SIMILARITY: Contains 2 CCHC-type zinc fingers.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; K03455; AAB50258.1; -  
 DR PDB; 1B6J; 11-DEC-00.  
 DR PDB; 1B6J; 21-APR-00.  
 DR PDB; 1NCP; 31-OCT-93.  
 DR PDB; 1TM; 12-NOV-96.  
 DR HIV; K03455; GAGSHXB2.  
 DR InterPro; IPR000721; Gag\_p24.  
 DR InterPro; IPR008916; Retrov\_capsid\_C.  
 DR InterPro; IPR008919; Retrov\_capsid\_N.  
 DR InterPro; IPR000071; Retrov\_p17.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00540; Gag\_p17; 1.  
 DR Pfam; PF00607; Gag\_p24; 1.  
 DR Pfam; PF00098; Zf\_CCHC; 2.  
 DR PRINTS; PR00939; C2HCZFINGER.  
 DR PRINTS; PR00234; HIVMATRIX.  
 DR SMART; SM00343; Znf\_CCHC; 2.  
 DR PROSITE; PS50158; Zf\_CCHC; 2.  
 DR AIDS; Core protein; Polypeptide; Myristate; Phosphorylation;  
 DR Zinc-finger; Repeat; 3D-structure; Lipoprotein.  
 DR INIT MET 0  
 DR CHAIN 1 131 CORE PROTEIN P17 (MATRIX PROTEIN).  
 DR CHAIN 132 362 CORE PROTEIN P24 (CORE ANTIGEN).  
 DR CHAIN 363 376 CORE PROTEIN P2.  
 DR CHAIN 377 431 CORE PROTEIN P7 (NUCLEOCAPSID PROTEIN).  
 DR CHAIN 432 447 CORE PROTEIN P1.  
 DR CHAIN 448 499 CORE PROTEIN P6.  
 DR ZN\_FING 389 406 CCHC-TYPE 1.  
 DR ZN\_FING 410 427 CCHC-TYPE 2.  
 DR LIPID 1 N-myristoyl glycine (in host) (By similarity).  
 DR SQ SEQUENCE 499 AA; 55798 MW; 774C384D6ACB108 CRC64;

Query Match 77.3%; Score 34; DB 1; Length 499;  
 Best Local Similarity 62.5%; Pred. No. 1.9e+02;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPSPPEE 8  
 Db 453 EPTAPPEE 460

RESULT 29  
 GAG\_HV1J3 STANDARD; PRT; 499 AA.  
 AC P1234;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE GAG polyprotein [contains: Core proteins P17, P24, P2, P7, P1, P6].  
 GN GAG.  
 OS Human immunodeficiency virus type 1 (HIV-1 isolate) (HIV-1).  
 OC Viruses; Retrov. Viruses; Retroviridae; Lentivirus.  
 NCBI\_TaxID=11694;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=89352108; PubMed=2669897;  
 RA Komiyama N., Hattori N., Inoue J., Sakuma S., Kurimura T., Yoshida M.;  
 RT "Nucleotide sequences of gag and env genes of a Japanese isolate of  
 RT HIV-1 and their expression in bacteria."  
 RT AIDS Res. Hum. Retroviruses 5:411-419(1989).  
 CC -1- FUNCTION: PERFORMS HIGHLY COMPLEX ORCHESTRATED TASKS DURING THE  
 CC ASSEMBLY, BUDDING, MATURATION, AND INFECTION STAGES OF THE VIRAL  
 CC REPLICATION CYCLE. DURING VIRAL ASSEMBLY, THE PROTEINS FORM  
 CC MEMBRANE ASSOCIATIONS AND SELF-ASSOCIATIONS THAT IMMEDIATELY  
 CC RESULT IN BUDDING OF AN IMMATURE VIRION FROM THE INFECTED CELL.  
 CC GAG PRECURSORS ALSO FUNCTION DURING VIRAL ASSEMBLY TO SELECTIVELY  
 CC BIND AND PACKAGE TWO PLUS STRANDS OF GENOMIC RNA.

CC -1- P17: The p24 protein is phosphorylated.  
 CC -1- SIMILARITY: Contains 2 CCHC-type zinc fingers.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M21137; AAB03522.1; -  
 DR HSSP; P05888; IAA.  
 DR HIV; M21137; GAGSHJ3.  
 DR InterPro; IPR000721; Gag\_p24.  
 DR InterPro; IPR008916; Retrov\_capsid\_C.  
 DR InterPro; IPR008919; Retrov\_capsid\_N.  
 DR InterPro; IPR000071; Retrov\_p17.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00540; Gag\_p17; 1.  
 DR Pfam; PF00607; Gag\_p24; 1.  
 DR Pfam; PF00098; Zf\_CCHC; 2.  
 DR PRINTS; PR00939; C2HCZFINGER.  
 DR PRINTS; PR00234; HIVMATRIX.  
 DR SMART; SM00343; Znf\_CCHC; 2.  
 DR PROSITE; PS50158; Zf\_CCHC; 2.  
 DR AIDS; Core protein; Polypeptide; Myristate; Phosphorylation;  
 DR Zinc-finger; Repeat; Lipoprotein.  
 DR INIT MET 0  
 DR CHAIN 1 131 CORE PROTEIN P17 (MATRIX PROTEIN).  
 DR CHAIN 132 362 CORE PROTEIN P24 (CORE ANTIGEN).  
 DR CHAIN 363 376 CORE PROTEIN P2.  
 DR CHAIN 377 431 CORE PROTEIN P7 (NUCLEOCAPSID PROTEIN).  
 DR CHAIN 432 447 CORE PROTEIN P1.  
 DR CHAIN 448 499 CORE PROTEIN P6.  
 DR ZN\_FING 389 406 CCHC-TYPE 1.  
 DR ZN\_FING 410 427 CCHC-TYPE 2.  
 DR LIPID 1 N-myristoyl glycine (in host) (By similarity).  
 DR SQ SEQUENCE 499 AA; 55725 MW; 52E2812D4424ABE9 CRC64;

Query Match 77.3%; Score 34; DB 1; Length 499;  
 Best Local Similarity 62.5%; Pred. No. 1.9e+02;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPSPPEE 8  
 Db 453 EPTAPPEE 460

RESULT 30  
 GAG\_HV1LW STANDARD; PRT; 499 AA.  
 AC Q70622;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE GAG polyprotein [contains: Core proteins P17, P24, P2, P7, P1, P6].  
 GN GAG.  
 OS Human immunodeficiency virus type 1 (HIV-1 isolate) (HIV-1).  
 OC Viruses; Retrov. Viruses; Retroviridae; Lentivirus.  
 NCBI\_TaxID=82834;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=95127297; PubMed=7826699;  
 RA Raitz M.S., Jr., Hall L., Robert-Guroff M., Lautenberger J., Hahn B.M.,  
 RA Shaw G.M., Kong L.I., Weiss S.H., Waters D., Gallo R.C., Blattner W.;  
 RT "Viral variability and serum antibody response in a laboratory worker  
 RT infected with HIV type 1 (HIV type 11B)."  
 RT AIDS Res. Hum. Retroviruses 10:1143-1155(1994).  
 CC -1- FUNCTION: PERFORMS HIGHLY COMPLEX ORCHESTRATED TASKS DURING THE  
 CC ASSEMBLY, BUDDING, MATURATION, AND INFECTION STAGES OF THE VIRAL  
 CC REPLICATION CYCLE. DURING VIRAL ASSEMBLY, THE PROTEINS FORM

```
CC MEMBRANE ASSOCIATIONS AND SELF-ASSOCIATIONS THAT ULTIMATELY
CC RESULT IN BUDDING OF AN IMMATURE VIRION FROM THE INFECTED CELL.
CC GAG PRECURSORS ALSO FUNCTION DURING VIRAL ASSEMBLY TO SELECTIVELY
CC BIND AND PACKAGE TWO PLUS STRANDS OF GENOMIC RNA.
CC -1- PTM: The p24 protein is phosphorylated.
CC -1- SIMILARITY: Contains 2 CCHC-type zinc fingers.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U12055; AAA76686.1; -.
DR PDB: 1HVN; 31-JAN-94.
DR PDB: 1HVO; 31-JAN-94.
DR PDB: 2ZNF; 15-JUL-92.
DR InterPro: IPR000721; Gag_p24.
DR InterPro: IPR008916; Retrov_capsid_C.
DR InterPro: IPR008919; Retrov_capsid_N.
DR InterPro: IPR000071; Retrov_p17.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00540; Gag_p17; 1.
DR Pfam: PF00607; Gag_p24; 1.
DR Pfam: PF00098; Zf_CCHC; 2.
DR PRINTS: PR00939; C2HCZNFINGER.
DR PRINTS: PR00234; HIVMATTRIX.
DR SMART: SM00343; Znf_C2HC; 2.
DR PROSITE: PS50158; ZF_CCHC; 2.
DR AIDS: Core protein; Polyprotein; Myristate; Phosphorylation;
DR Zinc-finger; Repeat; 3D-structure; Lipoprotein.
DR INIT MET 0
DR CHAIN 1 131 CORE PROTEIN P17 (MATRIX PROTEIN).
DR CHAIN 132 362 CORE PROTEIN P24 (CORE ANTIGEN).
DR CHAIN 363 376 CORE PROTEIN P2.
DR CHAIN 377 431 CORE PROTEIN P7 (NUCLEOCAPSID PROTEIN).
DR CHAIN 432 447 CORE PROTEIN P1.
DR CHAIN 448 499 CORE PROTEIN P6.
DR ZN_FING 389 406 CCHC-TYPE 1.
DR ZN_FING 410 427 CCHC-TYPE 2.
DR LIPID 1 1 N-myristoyl glycine (in host) (By
DR FT similarity).
DR SQ SEQUENCE 499 AA; 55772 MW; AA2F683546EDC0A9 CRC64;
Query Match 77.3%; Score 34; DB 1; Length 499;
Best Local Similarity 62.5%; Pred. No. 1.9e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 KPSSPPE 8
Db 453 EPTAPPE 460
RESULT 31
GAG_HVINS STANDARD; PRT; 499 AA.
AC P12493;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE GAG polyprotein [contains: Core proteins p17, p24, p7, p1, p6].
GN GAG.
OS Human immunodeficiency virus type 1 (New York-5 isolate) (HIV-1).
OC Viruses; Retrovirda; Retroviridae; Lentivirus.
OX NCB1_TaxID=11698;
RN [1]
RP SEQUENCE FROM N.A. (CLONE PNL4-3).
RA Buckler C.E., Buckler-White A.J., Willey R.L., McCoy J.;
RT Submitted (JUN-1988) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PERFORMS HIGHLY COMPLEX ORCHESTRATED TASKS DURING THE
CC ASSEMBLY, BUDDING, MATURATION, AND INFECTION STAGES OF THE VIRAL
```

```
CC REPLICATION CYCLE. DURING VIRAL ASSEMBLY, THE PROTEINS FORM
CC MEMBRANE ASSOCIATIONS AND SELF-ASSOCIATIONS THAT ULTIMATELY
CC RESULT IN BUDDING OF AN IMMATURE VIRION FROM THE INFECTED CELL.
CC GAG PRECURSORS ALSO FUNCTION DURING VIRAL ASSEMBLY TO SELECTIVELY
CC BIND AND PACKAGE TWO PLUS STRANDS OF GENOMIC RNA.
CC -1- PTM: The p24 protein is phosphorylated.
CC -1- SIMILARITY: Contains 2 CCHC-type zinc fingers.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19921; AAA44987.1; -.
DR PDB: 1A43; 09-FEB-99.
DR PDB: 1A80; 28-OCT-98.
DR PDB: 1AFV; 20-AUG-97.
DR PDB: 1AK4; 15-OCT-97.
DR PDB: 1AUM; 14-JAN-98.
DR PDB: 1BAJ; 18-NOV-98.
DR PDB: 1GDS; 30-DEC-96.
DR PDB: 1GDI; 30-DEC-96.
DR PDB: 1GDI; 30-DEC-96.
DR PDB: 1GDI; 30-DEC-96.
DR PDB: 1GDI; 30-DEC-96.
DR PDB: 1GDI; 30-DEC-96.
DR HIV: M19921; GAGSLN43.
DR InterPro: IPR000721; Gag_p24.
DR InterPro: IPR008916; Retrov_capsid_C.
DR InterPro: IPR008919; Retrov_capsid_N.
DR InterPro: IPR000071; Retrov_p17.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00540; Gag_p17; 1.
DR Pfam: PF00607; Gag_p24; 1.
DR Pfam: PF00098; Zf_CCHC; 2.
DR PRINTS: PR00939; C2HCZNFINGER.
DR PRINTS: PR00234; HIVMATTRIX.
DR SMART: SM00343; Znf_C2HC; 2.
DR PROSITE: PS50158; ZF_CCHC; 2.
DR AIDS: Core protein; Polyprotein; Myristate; Phosphorylation;
DR Zinc-finger; Repeat; 3D-structure; Lipoprotein.
DR INIT MET 0
DR CHAIN 1 131 CORE PROTEIN P17 (MATRIX PROTEIN).
DR CHAIN 132 362 CORE PROTEIN P24 (CORE ANTIGEN).
DR CHAIN 363 376 CORE PROTEIN P2.
DR CHAIN 377 431 CORE PROTEIN P7 (NUCLEOCAPSID PROTEIN).
DR CHAIN 432 447 CORE PROTEIN P1.
DR CHAIN 448 499 CORE PROTEIN P6.
DR ZN_FING 389 406 CCHC-TYPE 1.
DR ZN_FING 410 427 CCHC-TYPE 2.
DR LIPID 1 1 N-myristoyl glycine (in host) (By
DR FT similarity).
DR SQ SEQUENCE 499 AA; 55687 MW; C8EC1302FE2C1E2 CRC64;
Query Match 77.3%; Score 34; DB 1; Length 499;
Best Local Similarity 62.5%; Pred. No. 1.9e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 KPSSPPE 8
Db 453 EPTAPPE 460
RESULT 32
GAG_HVIRH STANDARD; PRT; 500 AA.
AC P05890;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE GAG polyprotein [contains: Core proteins p17, p24, p7, p1, p6].
GN GAG.
```



OS Human immunodeficiency virus type 1 (RF/HAT isolate) (HIV-1).  
 OC Viruses; Retroviral viruses; Retroviridae; Lentiviruses.  
 OX NCBI\_TaxID=11701;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Starcich B.R., Hahn B.H., Shaw G.M., McNeely P.D., Modrow S.,  
 RA Wolf H., Parks E.S., Parks W.P., Josephs S.F., Gallo R.C.,  
 RA Wong-Staal F.;  
 RL Submitted (XX-1987) to the HIV data bank.  
 CC -1- FUNCTION: PERFORMS HIGHLY COMPLEX ORCHESTRATED TASKS DURING THE  
 CC ASSEMBLY, BUDDING, MATURATION, AND INFECTION STAGES OF THE VIRAL  
 CC REPLICATION CYCLE. DURING VIRAL ASSEMBLY, THE PROTEINS FORM  
 CC MEMBRANE ASSOCIATIONS AND SELF-ASSOCIATIONS THAT ULTIMATELY  
 CC RESULT IN BUDDING OF AN IMMATURE VIRION FROM THE INFECTED CELL.  
 CC GAG PRECURSORS ALSO FUNCTION DURING VIRAL ASSEMBLY TO SELECTIVELY  
 CC BIND AND PACKAGE TWO PLUS STRANDS OF GENOMIC RNA.  
 CC -1- PTM: The p24 protein is phosphorylated.  
 CC -1- SIMILARITY: Contains 2 CCHC-type zinc fingers.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
 CC or send an email to [license@1sb-sib.ch](mailto:license@1sb-sib.ch)).  
 CC -----  
 DR EMBL, M17451; AAA45052.1; --  
 DR HSSP, P05888; 1AAF.  
 DR HIV, M17451; GAGSFR.  
 DR InterPro: IPR000721; Gag\_p24.  
 DR InterPro: IPR008916; Retrov\_capsid\_C.  
 DR InterPro: IPR008919; Retrov\_capsid\_N.  
 DR InterPro: IPR001878; Znf\_CCHC.  
 DR InterPro: IPR000071; Retrov\_p17.  
 DR Pfam: PF00540; Gag\_p17; 1.  
 DR Pfam: PF00607; Gag\_p24; 1.  
 DR Pfam: PF00098; Zf-CCHC; 2.  
 DR PRINTS, PR00939; CCHCNFINGER.  
 DR PRINTS, PR00234; HIVMATRIX.  
 DR SMART, SM00343; Znf\_CCHC; 2.  
 DR PROSITE, PSS0158; ZF\_CCHC; 2.  
 DR AIDS; Core protein; Polyprotein; Myristate; Phosphorylation;  
 DR ZINC-finger; Repeat; Lipoprotein.  
 DR INIT\_MET 0  
 FT CHAIN 1 131 CORE PROTEIN P17 (MATRIX PROTEIN).  
 FT CHAIN 132 362 CORE PROTEIN P24 (CORE ANTIGEN).  
 FT CHAIN 363 376 CORE PROTEIN P2.  
 FT CHAIN 377 433 CORE PROTEIN P7 (NUCLEOCAPSID PROTEIN).  
 FT CHAIN 434 449 CORE PROTEIN P1.  
 FT CHAIN 450 500 CORE PROTEIN P6.  
 FT ZN\_FING 389 406 CCHC-TYPE 1.  
 FT ZN\_FING 406 427 CCHC-TYPE 2.  
 FT ZN\_FING 410 427 N-myristoyl glycine (in host) (By  
 FT LIPID 1 similarity).  
 SQ SEQUENCE 500 AA; 55825 MW; 44AA0CB5CD4EF7B CRC64;  
 QY Query Match 77.3%; Score 34; DB 1; Length 500;  
 Db Best Local Similarity 62.5%; Pred. No. 1.9e+02;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KPSSPPE 8  
 Db 455 EPTAPPE 462  
 RESULT 33  
 ID \_GAG\_HV1A2 STANDARD; PRT; 501 AA.  
 AC P03349;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE GAG polyprotein [contains: Core proteins P17, P24, P2, P7, P1, P6].  
 GN GAG.  
 OS Human immunodeficiency virus type 1 (ARV2/SF2 isolate) (HIV-1).  
 OC Viruses; Retroviral viruses; Retroviridae; Lentiviruses.  
 OX NCBI\_TaxID=11685;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=85090453; PubMed=2578227;  
 RA Sanchez-Pescador R., Power M.D., Barr P.J., Steiner K.S.,  
 RA Stempien M.W., Brown-Shimer S.L., Gee W.W., Renard A., Randolph A.,  
 RA Levy J.A., Dina D., Luciw P.A.;  
 RT "Nucleotide sequence and expression of an AIDS-associated retrovirus  
 RT (ARV-2)";  
 RL Science 227:484-492 (1985).  
 CC -1- FUNCTION: PERFORMS HIGHLY COMPLEX ORCHESTRATED TASKS DURING THE  
 CC ASSEMBLY, BUDDING, MATURATION, AND INFECTION STAGES OF THE VIRAL  
 CC REPLICATION CYCLE. DURING VIRAL ASSEMBLY, THE PROTEINS FORM  
 CC MEMBRANE ASSOCIATIONS AND SELF-ASSOCIATIONS THAT ULTIMATELY  
 CC RESULT IN BUDDING OF AN IMMATURE VIRION FROM THE INFECTED CELL.  
 CC GAG PRECURSORS ALSO FUNCTION DURING VIRAL ASSEMBLY TO SELECTIVELY  
 CC BIND AND PACKAGE TWO PLUS STRANDS OF GENOMIC RNA.  
 CC -1- PTM: The p24 protein is phosphorylated.  
 CC -1- SIMILARITY: Contains 2 CCHC-type zinc fingers.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
 CC or send an email to [license@1sb-sib.ch](mailto:license@1sb-sib.ch)).  
 CC -----  
 DR EMBL, K02007; AAB5875.1; --  
 DR PIR, A03947; FOVM2.  
 DR HSSP, P05888; 1AAF.  
 DR HIV, K02007; GAGSFR2.  
 DR InterPro: IPR000721; Gag\_p24.  
 DR InterPro: IPR008916; Retrov\_capsid\_C.  
 DR InterPro: IPR008919; Retrov\_capsid\_N.  
 DR InterPro: IPR000071; Retrov\_p17.  
 DR InterPro: IPR001878; Znf\_CCHC.  
 DR Pfam: PF00540; Gag\_p17; 1.  
 DR Pfam: PF00607; Gag\_p24; 1.  
 DR Pfam: PF00098; Zf-CCHC; 2.  
 DR PRINTS, PR00939; CCHCNFINGER.  
 DR PRINTS, PR00234; HIVMATRIX.  
 DR SMART, SM00343; Znf\_CCHC; 2.  
 DR PROSITE, PSS0158; ZF\_CCHC; 2.  
 DR AIDS; Core protein; Polyprotein; Myristate; Phosphorylation;  
 DR ZINC-finger; Repeat; Lipoprotein.  
 DR INIT\_MET 0  
 FT CHAIN 1 133 CORE PROTEIN P17 (MATRIX PROTEIN).  
 FT CHAIN 134 364 CORE PROTEIN P24 (CORE ANTIGEN).  
 FT CHAIN 365 378 CORE PROTEIN P2.  
 FT CHAIN 379 433 CORE PROTEIN P7 (NUCLEOCAPSID PROTEIN).  
 FT CHAIN 434 449 CORE PROTEIN P1.  
 FT CHAIN 450 501 CORE PROTEIN P6.  
 FT ZN\_FING 391 408 CCHC-TYPE 1.  
 FT ZN\_FING 408 429 CCHC-TYPE 2.  
 FT ZN\_FING 412 429 N-myristoyl glycine (in host) (By  
 FT LIPID 1 similarity).  
 SQ SEQUENCE 501 AA; 55935 MW; 8F36C928D5DBA45 CRC64;  
 QY Query Match 77.3%; Score 34; DB 1; Length 501;  
 Db Best Local Similarity 62.5%; Pred. No. 1.9e+02;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KPSSPPE 8  
 Db 455 EPTAPPE 462  
 RESULT 34

GAG\_HV1JR  
ID GAG\_HV1JR STANDARD; PRT; 503 AA.  
AC P20873;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE GAG polypeptide [Contains: Core proteins P17, P24, P2, P7, P1, P6].  
GN GAG.  
OS Human immunodeficiency virus type 1 (JRCSF isolate) (HIV-1).  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11696;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Koyanagi S., Chen I.S.Y.,  
RL Submitted (DEC-1988) to the HIV data bank.  
CC -1- FUNCTION: PERFORMS HIGHLY COMPLEX ORCHESTRATED TASKS DURING THE  
ASSEMBLY, BUDDING, MATURATION, AND INJECTION STAGES OF THE VIRAL  
REPLICATION CYCLE. DURING VIRAL ASSEMBLY, THE PROTEINS FORM  
MEMBRANE ASSOCIATIONS AND SELF-ASSOCIATIONS THAT ULTIMATELY  
RESULT IN BUDDING OF AN IMMATURE VIRION FROM THE INFECTED CELL.  
CC GAG PRECURSORS ALSO FUNCTION DURING VIRAL ASSEMBLY TO SELECTIVELY  
CC BIND AND PACKAGE TWO PLUS STRANDS OF GENOMIC RNA.  
CC -1- PTM: The p24 protein is phosphorylated.  
CC -1- SIMILARITY: Contains 2 CCHC-type zinc fingers.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).  
CC -----  
DR EMBL; M38429; AAB03744.1; -.  
DR PDB; 1KJ4; 06-MAR-02.  
DR PDB; 1KJ7; 06-MAR-02.  
DR PDB; 1KJF; 06-MAR-02.  
DR HIV; M38429; GAGSURCSF.  
DR InterPro: IPR000721; Gag\_P24.  
DR InterPro: IPR008916; Retrov\_capsid\_C.  
DR InterPro: IPR008919; Retrov\_capsid\_N.  
DR InterPro: IPR000071; Retrov\_P17.  
DR InterPro: IPR001878; Znf\_CCHC.  
DR Pfam; PF00540; Gag\_P17; 1.  
DR Pfam; PF00607; Gag\_P24; 1.  
DR Pfam; PF00098; zf-CCHC; 2.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR PRINTS; PR00234; HIVMATRIX.  
DR SMART; SM00343; Znf\_CCHC; 2.  
DR PROSITE; PS0158; ZF\_CCHC; 2.  
KW AIDS; Core protein; Polypeptide; Myristate; Phosphorylation;  
KW zinc-finger; Repeat; 3D-structure; Lipoprotein.  
FT INIT MET 0  
FT CHAIN 1 131 CORE PROTEIN P17 (MATRIX PROTEIN).  
FT CHAIN 132 362 CORE PROTEIN P24 (CORE ANTIGEN).  
FT CHAIN 363 376 CORE PROTEIN P2.  
FT CHAIN 377 431 CORE PROTEIN P7 (NUCLEOCAPSID PROTEIN).  
FT CHAIN 432 447 CORE PROTEIN P1.  
FT CHAIN 448 503 CORE PROTEIN P6.  
FT ZN\_FING 389 406 CCHC-TYPE 1.  
FT ZN\_FING 410 427 CCHC-TYPE 2.  
FT LIPID 1 1 N-myristoyl glycine (in host) (By  
similarity).  
SQ SEQUENCE 503 AA; 56352 MM; F0B0471CC09B42BD CRC64;  
Query Match 77.3%; Score 34; DB 1; Length 503;  
Best Local Similarity 62.5%; Pred. No. 1.9e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

RESULT 35  
GAG\_HV1JM  
ID GAG\_HV1JM STANDARD; PRT; 506 AA.  
AC P05888;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE GAG polypeptide [Contains: Core proteins P17, P24, P2, P7, P1, P6].  
GN GAG.  
OS Human immunodeficiency virus type 1 (MN isolate) (HIV-1).  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11696;  
RN [1]  
RP SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.  
RX MEDLINE=92194415; PubMed=1548743;  
RA Henderson L.E., Bowers M.A., Sowder R.C. II, Serbyn S.A.,  
RA Johnson D.G., Bees J.W., Jr., Arthur L.O., Bryant D.K., Fenselau C.,  
RT "Gag proteins of the highly replicative MN strain of human  
immunodeficiency virus type 1: posttranslational modifications,  
proteolytic processings, and complete amino acid sequences.";  
RT J. Virol. 66:1856-1865(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88219542; PubMed=3369091;  
RA Gargo C., Guo H.-G., Franchini G., Aldovini A., Collalti E.,  
RA Farrell K., Wong-Staal F., Gallo R.C., Reitz M.S. Jr.,  
RT "Envelope sequences of two new United States HIV-1 isolates.";  
RT Virology 164:531-536(1988).  
RN [3]  
RP STRUCTURE BY NMR OF 380-434.  
RX MEDLINE=93278285; PubMed=1304355;  
RA Summers M.F., Henderson L.E., Chance M.R., Bees J.W. Jr., South T.L.,  
RA Blake P.R., Sagl I., Perez-Alvarado G., Sowder R.C. III, Hare D.R.,  
RA Arthur L.O.,  
RT "Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of  
intact viruses and the solution-state structure of the nucleocapsid  
protein from HIV-1.";  
RT Protein Sci. 1:563-574(1992).  
CC -1- FUNCTION: PERFORMS HIGHLY COMPLEX ORCHESTRATED TASKS DURING THE  
ASSEMBLY, BUDDING, MATURATION, AND INJECTION STAGES OF THE VIRAL  
REPLICATION CYCLE. DURING VIRAL ASSEMBLY, THE PROTEINS FORM  
MEMBRANE ASSOCIATIONS AND SELF-ASSOCIATIONS THAT ULTIMATELY  
RESULT IN BUDDING OF AN IMMATURE VIRION FROM THE INFECTED CELL.  
CC GAG PRECURSORS ALSO FUNCTION DURING VIRAL ASSEMBLY TO SELECTIVELY  
CC BIND AND PACKAGE TWO PLUS STRANDS OF GENOMIC RNA.  
CC -1- PTM: The p24 protein is phosphorylated.  
CC -1- MISCELLANEOUS: THE MN ISOLATE WAS TAKEN FROM A PEDIATRIC AIDS  
PATIENT IN 1984.  
CC -1- SIMILARITY: Contains 2 CCHC-type zinc fingers.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).  
CC -----  
DR EMBL; M17449; AAA44853.1; -.  
DR PIR; A38068; A38068.  
DR PDB; 1AAF; 31-JAN-94.  
DR HIV; M17449; GAGSMN.  
DR InterPro: IPR000721; Gag\_P24.  
DR InterPro: IPR008916; Retrov\_capsid\_C.  
DR InterPro: IPR008919; Retrov\_capsid\_N.  
DR InterPro: IPR001878; Znf\_CCHC.  
DR Pfam; PF00540; Gag\_P17; 1.  
DR Pfam; PF00607; Gag\_P24; 1.  
DR Pfam; PF00098; zf-CCHC; 2.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR PRINTS; PR00234; HIVMATRIX.

```

DR SMART; SM00343; ZNF C2HC; 2.
DR PROSITE; PS0158; ZF_C2HC; 2.
KW AIDS; Core protein; Polypeptide; Myristate; Phosphorylation;
KW Zinc-finger; 3D-structure; Repeat; Lipoprotein.
FT CHAIN 1 134 CORE PROTEIN P17 (MATRIX ANTIGEN).
FT INIT MET 0 CORE PROTEIN P24 (CORE ANTIGEN).
FT CHAIN 135 365 CORE PROTEIN P2.
FT CHAIN 366 379 CORE PROTEIN P7 (NUCLEOCAPSID PROTEIN).
FT CHAIN 380 434 CORE PROTEIN P1.
FT CHAIN 435 450 CORE PROTEIN P6.
FT CHAIN 451 506 CORE PROTEIN P1.
FT ZN_FING 392 409 CCHC-TYPE 1.
FT ZN_FING 413 430 CCHC-TYPE 2.
FT LIPID 1 1 N-myristoyl glycine (in host).
FT VARIANT 34 34 I -> V.
FT VARIANT 45 45 R -> L OR S OR N.
FT VARIANT 74 74 K -> E.
FT VARIANT 92 92 K -> N (IN REF. 2).
FT CONFLICT 17 17 Q -> E (IN REF. 2).
FT CONFLICT 141 141 A -> V (IN REF. 2).
FT CONFLICT 220 226 A -> T (IN REF. 2).
FT CONFLICT 318 319 WM -> RT (IN REF. 2).
FT CONFLICT 447 448 PG -> R (IN REF. 2).
FT TURN 395 397
FT HELIX 404 406
FT TURN 416 418
FT HELIX 425 427
SQ SEQUENCE 506 AA; 56629 MW; AC6F3CEB691C4726 CRC64;

Query Match 77.3%; Score 34; DB 1; Length 506;
Best Local Similarity 62.5%; Pred. No. 1.9e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSPPEE 8
Db 456 EPTAPPEE 463

RESULT 36
GAG_HV1B1 STANDARD; PRT; 511 AA.
AC P03347;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE GAG polyprotein [Contains: Core proteins P17, P24, P2, P7, P1, P6].
GN GAG.
OS Human immunodeficiency virus type 1 (BH10 isolate) (HIV-1).
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_Taxid=11678;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8511123; PubMed=2578615;
RA Ratner L., Haseltine W., Patarca R., Livak K.J., Starcich B.R.,
RA Josephs S.F., Doran E.R., Ratafski J.A., Whitehorn E.A.,
RA Baumeister K., Ivanoff L., Petteway S.R. Jr., Pearson M.L.,
RA Lautenberger J.A., Papas T.S., Ghayab J., Chang N.T., Gallo R.C.,
RA Wong-Staal F.;
RA "Complete nucleotide sequence of the AIDS virus, HTLV-III.";
RT Nature 313:277-284(1985).
CC -1- FUNCTION: PERFORMS HIGHLY COMPLEX ORCHESTRATED TASKS DURING THE
CC ASSEMBLY, BUDDING, MATURATION, AND INFECTION STAGES OF THE VIRAL
CC REPLICATION CYCLE. DURING VIRAL ASSEMBLY, THE PROTEINS FORM
CC MEMBRANE ASSOCIATIONS AND SELF-ASSOCIATIONS THAT ULTIMATELY
CC RESULT IN BUDDING OF AN IMMATURE VIRION FROM THE INFECTED CELL.
CC GAG PRECURSORS ALSO FUNCTION DURING VIRAL ASSEMBLY TO SELECTIVELY
CC BIND AND PACKAGE TWO PLUS STRANDS OF GENOMIC RNA.
CC -1- PTM: The p24 protein is phosphorylated.
CC -1- SIMILARITY: Contains 2 CCHC-type zinc fingers.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M15654; AAA44201.1; -.
DR PIR; A03945; FOVMH3.
DR HSP; P05888; 1A8F.
DR HIV; M15654; GAGSH102.
DR InterPro; IPR000721; Gag_p24.
DR InterPro; IPR008916; Retrov_capsid_C.
DR InterPro; IPR008919; Retrov_capsid_N.
DR InterPro; IPR000071; Retrovir_p17.
DR InterPro; IPR001878; Znf C2HC.
DR Pfam; PF00540; Gag_p17; 1.
DR Pfam; PF00607; Gag_p24; 1.
DR Pfam; PF00098; ZF_C2HC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR PRINTS; PR00234; HIVMATRIX.
DR SMART; SM00343; ZNF C2HC; 2.
DR PROSITE; PS0158; ZF_C2HC; 2.
KW AIDS; Core protein; Polypeptide; Myristate; Phosphorylation;
KW Zinc-finger; Repeat; Lipoprotein.
FT CHAIN 1 131 CORE PROTEIN P17 (MATRIX PROTEIN).
FT INIT MET 0 CORE PROTEIN P24 (CORE ANTIGEN).
FT CHAIN 132 362 CORE PROTEIN P2.
FT CHAIN 363 376 CORE PROTEIN P7 (NUCLEOCAPSID PROTEIN).
FT CHAIN 377 431 CORE PROTEIN P1.
FT CHAIN 432 447 CORE PROTEIN P6.
FT CHAIN 448 511 CORE PROTEIN P6.
FT ZN_FING 389 406 CCHC-TYPE 1.
FT ZN_FING 410 427 CCHC-TYPE 2.
FT LIPID 1 1 N-myristoyl glycine (in host) (By
FT similarity).
SQ SEQUENCE 511 AA; 57125 MW; 97C95188EB0D458 CRC64;

Query Match 77.3%; Score 34; DB 1; Length 511;
Best Local Similarity 62.5%; Pred. No. 1.9e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSPPEE 8
Db 465 EPTAPPEE 472

RESULT 37
GAG_HV1B5 STANDARD; PRT; 511 AA.
AC P04593;
DT 11-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE GAG polyprotein [Contains: Core proteins P17, P24, P2, P7, P1, P6].
GN GAG.
OS Human immunodeficiency virus type 1 (BH5 isolate) (HIV-1).
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_Taxid=11682;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8511123; PubMed=2578615;
RA Ratner L., Haseltine W., Patarca R., Livak K.J., Starcich B.R.,
RA Josephs S.F., Doran E.R., Ratafski J.A., Whitehorn E.A.,
RA Baumeister K., Ivanoff L., Petteway S.R. Jr., Pearson M.L.,
RA Lautenberger J.A., Papas T.S., Ghayab J., Chang N.T., Gallo R.C.,
RA Wong-Staal F.;
RA "Complete nucleotide sequence of the AIDS virus, HTLV-III.";
RT Nature 313:277-284(1985).
CC -1- FUNCTION: PERFORMS HIGHLY COMPLEX ORCHESTRATED TASKS DURING THE
CC ASSEMBLY, BUDDING, MATURATION, AND INFECTION STAGES OF THE VIRAL
CC REPLICATION CYCLE. DURING VIRAL ASSEMBLY, THE PROTEINS FORM
CC MEMBRANE ASSOCIATIONS AND SELF-ASSOCIATIONS THAT ULTIMATELY
CC RESULT IN BUDDING OF AN IMMATURE VIRION FROM THE INFECTED CELL.

```

```

CC GAG PRECURSORS ALSO FUNCTION DURING VIRAL ASSEMBLY TO SELECTIVELY
CC BIND AND PACKAGE TWO PLUS STRANDS OF GENOMIC RNA.
CC -1- P2M: The p24 protein is phosphorylated.
CC -1- SIMILARITY: Contains 2 CCHC-type zinc fingers.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; K02012; AAA44652.1; -.
CC HSSP; P05888; 1AAF.
CC HIV; K02012; GAGSBH5.
CC InterPro; IPR000721; Gag_p24.
CC InterPro; IPR008916; Retrov_capsid_C.
CC InterPro; IPR008919; Retrov_capsid_N.
CC InterPro; IPR000071; Retrovir_p17.
CC InterPro; IPR001878; Znf_CCHC.
CC Pfam; PF00540; Gag_p17; 1.
CC Pfam; PF00607; Gag_p24; 1.
CC Pfam; PF00098; Zf-CCHC; 2.
CC PRINTS; PR00939; C2HCZNFINGER.
CC PRINTS; PR00234; HIVMATR1X.
CC SMART; SM00343; ZNF_CCHC; 2.
CC PROSITE; PS50158; ZF_CCHC; 2.
CC AIDS; Core protein; Polyprotein; Myristate; Phosphorylation;
CC Zinc-finger; Repeat; Lipoprotein.
CC INIT MET 0
CC CHAIN 1 131 CORE PROTEIN P17 (MATRIX PROTEIN).
CC CHAIN 132 362 CORE PROTEIN P24 (CORE ANTIGEN).
CC CHAIN 363 376 CORE PROTEIN P2.
CC CHAIN 377 431 CORE PROTEIN P7 (NUCLEOCAPSID PROTEIN).
CC CHAIN 432 447 CORE PROTEIN P1.
CC CHAIN 448 511 CORE PROTEIN P6.
CC ZN_FING 389 406 CCHC-TYPE 1.
CC ZN_FING 410 427 CCHC-TYPE 2.
CC LIPID 1 1 N-myristoyl glycine (in host) (By
CC similarity).
CC SEQUENCE 511 AA; 57107 MW; B0AE72D3CBBE7A80 CRC64;
-----
QY Query Match 77.3%; Score 34; DB 1; Length 511;
Best Local Similarity 62.5%; Pred. No. 1.9e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 KPSSPPE 8
Db 465 EPTAPPE 472
-----
RESULT 38
GAG_HV1BR STANDARD; PRT; 511 AA.
AC P03348;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE GAG polyprotein [Contains: Core proteins p17, p24, p2, p7, p1, p6].
GN GAG.
OS Human immunodeficiency virus type 1 (BRU isolate) (HIV-1).
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OX NCBI_Taxid=11686;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=5039333; PubMed=2981635;
RT Wain-Hobson S., Sonigo P., Danos O., Cole S., Allizon M.;
RL "Nucleotide sequence of the AIDS virus, LAV.";
Cell 40:9-17(1985).
RN [2]
RP REVISIONS TO 459-470.
RX MEDLINE=86245056; PubMed=2424612;

```

```

RA Allizon M., Wain-Hobson S., Montagnier L., Sonigo P.;
RT "Genetic variability of the AIDS virus: nucleotide sequence analysis
RT of two isolates from African patients.";
Cell 46:63-74(1986).
RL [1]
CC -1- FUNCTION: PERFORMS HIGHLY COMPLEX ORCHESTRATED TASKS DURING THE
CC ASSEMBLY, BUDDING, MATURATION, AND INFECTION STAGES OF THE VIRAL
CC REPLICATION CYCLE. DURING VIRAL ASSEMBLY, THE PROTEINS FORM
CC MEMBRANE ASSOCIATIONS AND SELF-ASSOCIATIONS THAT ULTIMATELY
CC RESULT IN BUDDING OF AN IMMATURE VIRION FROM THE INFECTED CELL.
CC GAG PRECURSORS ALSO FUNCTION DURING VIRAL ASSEMBLY TO SELECTIVELY
CC BIND AND PACKAGE TWO PLUS STRANDS OF GENOMIC RNA.
CC -1- P2M: The p24 protein is phosphorylated.
CC -1- SIMILARITY: Contains 2 CCHC-type zinc fingers.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; K02013; AAB59747.1; -.
CC HSSP; P05888; 1AAF.
CC HIV; K02013; GAGSERU.
CC InterPro; IPR000721; Gag_p24.
CC InterPro; IPR008916; Retrov_capsid_C.
CC InterPro; IPR008919; Retrov_capsid_N.
CC InterPro; IPR000071; Retrovir_p17.
CC InterPro; IPR001878; Znf_CCHC.
CC Pfam; PF00540; Gag_p17; 1.
CC Pfam; PF00607; Gag_p24; 1.
CC Pfam; PF00098; Zf-CCHC; 2.
CC PRINTS; PR00939; C2HCZNFINGER.
CC PRINTS; PR00234; HIVMATR1X.
CC SMART; SM00343; ZNF_CCHC; 2.
CC PROSITE; PS50158; ZF_CCHC; 2.
CC AIDS; Core protein; Polyprotein; Myristate; Phosphorylation;
CC Zinc-finger; Repeat; Lipoprotein.
CC INIT MET 0
CC CHAIN 1 131 CORE PROTEIN P17 (MATRIX PROTEIN).
CC CHAIN 132 362 CORE PROTEIN P24 (CORE ANTIGEN).
CC CHAIN 363 376 CORE PROTEIN P2.
CC CHAIN 377 431 CORE PROTEIN P7 (NUCLEOCAPSID PROTEIN).
CC CHAIN 432 447 CORE PROTEIN P1.
CC CHAIN 448 511 CORE PROTEIN P6.
CC ZN_FING 389 406 CCHC-TYPE 1.
CC ZN_FING 410 427 CCHC-TYPE 2.
CC LIPID 1 1 N-myristoyl glycine (in host) (By
CC similarity).
CC SEQUENCE 511 AA; 57095 MW; 55972701A8434673 CRC64;
-----
QY Query Match 77.3%; Score 34; DB 1; Length 511;
Best Local Similarity 62.5%; Pred. No. 1.9e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 KPSSPPE 8
Db 465 EPTAPPE 472
-----
RESULT 39
GAG_HV1PV STANDARD; PRT; 511 AA.
AC P03350;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE GAG polyprotein [Contains: Core proteins p17, p24, p2, p7, p1, p6].
GN GAG.
OS Human immunodeficiency virus type 1 (PV22 isolate) (HIV-1).
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OX NCBI_Taxid=11700;

```

[1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=8511157; PubMed=2982104;  
 RA Muesing M.A., Smith D.H., Cabralilla C.D., Benton C.V., Laeky L.A.,  
 RT "Nucleic acid structure and expression of the human  
 RL AIDS/lymphadenopathy retrovirus.";  
 RN Nature 313:450-458(1985).  
 [2]  
 RP REVISION.  
 RA Muesing M.A.;  
 RL Submitted (XXX-1987) to the HIV data bank.  
 CC -1- FUNCTION: PERFORMS HIGHLY COMPLEX ORCHESTRATED TASKS DURING THE  
 ASSEMBLY, BUDDING, MATURATION, AND INFECTION STAGES OF THE VIRAL  
 CC REPLICATION CYCLE. DURING VIRAL ASSEMBLY, THE PROTEINS FORM  
 CC MEMBRANE ASSOCIATIONS AND SELF-ASSOCIATIONS THAT ULTIMATELY  
 CC RESULT IN BUDDING OF AN IMMATURE VIRION FROM THE INFECTED CELL.  
 CC GAG PRECURSORS ALSO FUNCTION DURING VIRAL ASSEMBLY TO SELECTIVELY  
 CC BIND AND PACKAGE TWO PLUS STRANDS OF GENOMIC RNA.  
 CC -1- PTM: The p24 protein is phosphorylated.  
 CC -1- SIMILARITY: Contains 2 CCHC-type zinc fingers.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; K02083; AAB59866.1; -  
 CC EMBL; X01762; CAA25902.1; ALT\_SEQ.  
 CC PIR; A03948; FOVWVL.  
 CC HSSP; P05888; 1AAP.  
 DR HIV; K02083; GAGSPV22.  
 DR InterPro; IPR000721; Gag\_p24.  
 DR InterPro; IPR008916; Retrov\_capsid\_C.  
 DR InterPro; IPR008919; Retrov\_capsid\_N.  
 DR InterPro; IPR000071; Retrov\_p17.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00540; Gag\_p17; 1.  
 DR Pfam; PF00607; Gag\_p24; 1.  
 DR Pfam; PF00098; zf\_CCHC; 2.  
 DR PRINTS; PR00939; CCHCNFINGER.  
 DR SMART; SM00234; HIVMATRIX.  
 DR SMART; SM00343; Znf\_CCHC; 2.  
 DR PROSITE; PS50158; ZF\_CCHC; 2.  
 KM AIDS; Core protein; Polypeptide; Myristate; Phosphorylation;  
 KM Zinc-finger; Repeat; Lipoprotein.  
 FT INIT MET 0  
 FT CHAIN 1 131 BY SIMILARITY.  
 FT CHAIN 132 362 CORE PROTEIN P17 (MATRIX PROTEIN).  
 FT CHAIN 363 376 CORE PROTEIN P24 (CORE ANTIGEN).  
 FT CHAIN 377 431 CORE PROTEIN P2.  
 FT CHAIN 432 447 CORE PROTEIN P7 (NUCLEOCAPSID PROTEIN).  
 FT CHAIN 448 511 CORE PROTEIN P1.  
 FT CHAIN 511 511 CORE PROTEIN P6.  
 FT ZN\_FING 389 406 CCHC-TYPE 1.  
 FT ZN\_FING 410 427 CCHC-TYPE 2.  
 FT LIPID 1 1 N-myristoyl glycine (in host) (By  
 FT similarity).  
 SQ SEQUENCE 511 AA; 57139 MW; 397601C9403095F2 CRC64;

Query Match 77.3%; Score 34; DB 1; Length 511;  
 Best Local Similarity 62.5%; Pred. No. 1.9e+02;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSSPPEE 8  
 DB 465 EPTAPPE 472  
 RESULT 40  
 MIS\_RAT

ID MIS\_RAT STANDARD; PRT; 553 AA.  
 AC P49000;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Muellerian inhibiting factor precursor (MIS) (Anti-muellerian hormone)  
 DE (AMH) (Muellerian inhibiting substance).  
 GN AMH.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92241861; PubMed=1572639;  
 RA Haq C., Lee M.M., Tizard R., Wyck M., Demarinis J., Donahoe P.K.,  
 RT "Isolation of the rat gene for Mullerian inhibiting substance.";  
 RL Genomics 12:665-669(1992).  
 CC -1- FUNCTION: THIS GLYCOPROTEIN, PRODUCED BY THE SERTOLI CELLS OF THE  
 TESTIS, CAUSES REGRESSION OF THE MUELLERIAN DUCT. IT ALSO IS ABLE  
 CC TO INHIBIT THE GROWTH OF TUMORS DERIVED FROM TISSUES OF MUELLERIAN  
 CC DUCT ORIGIN.  
 CC -1- SUBUNIT: Homodimer; disulfide-linked.  
 CC -1- SIMILARITY: Belongs to the TGF-beta family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; S98336; AAB22104.1; -  
 CC PIR; A42499; A42499.  
 DR InterPro; IPR006799; AMH\_N.  
 DR InterPro; IPR002400; GF\_CysKnot.  
 DR InterPro; IPR001839; TGFb.  
 DR Pfam; PF04709; AMH\_N; 1.  
 DR Pfam; PF00019; TGF-beta; 1.  
 DR PRINTS; PR00438; GF\_CYSKNOT.  
 DR PRODOM; PD000357; TGFb; 1.  
 DR SMART; SM00204; TGFb; 1.  
 DR PROSITE; PS00250; TGF\_BETA\_1; 1.  
 KM Growth factor; Glycoprotein; Gonadal differentiation; Signal.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT PROPEP 22 553 POTENTIAL.  
 FT CHAIN 1 553 MUELLERIAN INHIBITING FACTOR.  
 FT DISULFID 455 519 BY SIMILARITY.  
 FT DISULFID 481 550 BY SIMILARITY.  
 FT DISULFID 485 552 BY SIMILARITY.  
 FT DISULFID 518 518 INTERCHAIN (BY SIMILARITY).  
 FT CARBOHYD 325 325 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 409 409 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 553 AA; 58888 MW; 75DAF3949A038A69 CRC64;

Query Match 77.3%; Score 34; DB 1; Length 553;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PSSPPE 7  
 DB 44 PSSPPE 49  
 RESULT 41  
 MIS\_MOUSE  
 ID MIS\_MOUSE STANDARD; PRT; 555 AA.  
 AC P27106;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

QY 2 PSSPPE 7  
 DB 44 PSSPPE 49  
 RESULT 41  
 MIS\_MOUSE  
 ID MIS\_MOUSE STANDARD; PRT; 555 AA.  
 AC P27106;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Muellerian inhibiting factor precursor (MIS) (Anti-muellerian hormone)  
 DE AMH (Muellerian inhibiting substance).  
 GN AMH.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxId=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129; TISSUE=Testis;  
 RX MEDLINE=92146272; PubMed=1782869;  
 RA Muensterberg A., Lovell-Badge R.;  
 RT "Expression of the mouse anti-muellerian hormone gene suggests a role  
 in both male and female sexual differentiation."; Development 113:613-624 (1991).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=129;  
 RA Dresser D.W., Jamin S., Atkins C.J., Guerrier D.;  
 RT "A GNP-like gene shares a bidirectional promoter with SAP62  
 immediately upstream of AMH.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1-42 FROM N.A.  
 RC STRAIN=129;  
 RX MEDLINE=96081226; PubMed=8541848;  
 RA Dresser D.W., Hacker A., Lovell-Badge R., Guerrier D.;  
 RT "The genes for a spliceosome protein (SAP62) and the anti-Muellerian  
 hormone (AMH) are contiguous.";  
 RL Hum. Mol. Genet. 4:1613-1618 (1995).  
 CC -1- FUNCTION: THIS GLYCOPROTEIN, PRODUCED BY THE SERTOLI CELLS OF THE  
 TESTIS, CAUSES REGRESSION OF THE MUELLERIAN DUCT. IT ALSO IS ABLE  
 TO INHIBIT THE GROWTH OF TUMORS DERIVED FROM TISSUES OF MUELLERIAN  
 DUCT ORIGIN.  
 CC -1- SUBUNIT: Homodimer; disulfide-linked.  
 CC -1- TISSUE SPECIFICITY: SERTOLI CELLS OF TESTIS, AND TESTES  
 JUST AFTER BIRTH, BUT ABSENT IN ADULT TESTES. IN FEMALE, AMH  
 IS EXPRESSED AFTER BIRTH IN THE GRANULOSA CELLS OF FOLLICULAR  
 MATURATION AND NOT ON THE AGE OF THE OVARY.  
 CC -1- MISCELLANEOUS: Although it does not compete with EGF for receptor  
 binding sites, MIS can inhibit the autophosphorylation of the EGF  
 receptor in vitro.  
 CC -1- SIMILARITY: Belongs to the TGF-beta family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X63240; CAA44912.1; -;  
 DR EMBL: X83733; CAC10450.1; -;  
 DR PIR: S20100; S20100.  
 DR MGD: MGI:88006; Amh.  
 DR InterPro: IPR006799; AMH N.  
 DR InterPro: IPR002400; GF\_Cysknoc.  
 DR InterPro: IPR001839; TGFb.  
 DR Pfam: PF04709; AMH\_N; 1.  
 DR Pfam: PF00019; TGF-beta; 1.  
 DR PRINTS: PR00438; GFCYSKNOC.  
 DR ProDom: PD000357; TGFb; 1.  
 DR SMART: SM00204; TGFb; 1.  
 DR PROSITE: PS00250; TGF\_BETA\_1; 1.  
 KW Growth factor; Glycoprotein; Gonadal differentiation; Signal.  
 FT SIGNAL 1 20  
 FT PROPEP 21 ?  
 FT CHAIN 2 555  
 FT DISULFID 457 521  
 FT DISULFID 483 552  
 FT DISULFID 487 554  
 FT BY SIMILARITY.  
 FT BY SIMILARITY.  
 FT BY SIMILARITY.

FT DISULFID 520 520 INTERCHAIN (BY SIMILARITY).  
 FT CARBOHYD 62 62 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 326 326 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 555 AA; 59778 MW; D6A3A20C50306E29 CRC64;  
 Query Match 77.3%; Score 34; DB 1; Length 555;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 45 PSSPPE 50  
 QY 2 PSSPPE 7  
 DB 45 PSSPPE 50  
 RESULT 42  
 NR42 HUMAN STANDARD; PRT; 598 AA.  
 AC P43354;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Orphan nuclear receptor NURR1 (immediate-early response protein NOT  
 DE (transcriptionally inducible nuclear receptor).  
 GN NR4A2 OR NURR1 OR T1NOR OR NOT.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Blood;  
 RX MEDLINE=95183071; PubMed=1877627;  
 RA Mages H.W., Rilke O., Bravo R., Senger G., Kroczeck R.A.;  
 RT "NOT, a human immediate-early response gene closely related to the  
 RT steroid/thyroid hormone receptor NAK1/TR3.";  
 RL Mol. Endocrinol. 8:1583-1591 (1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Fibroblast;  
 RX MEDLINE=99234104; PubMed=10216262;  
 RA Ichinose H., Ohye T., Suzuki T., Sumi-Ichinose C., Nomura T.,  
 RA Hagino Y., Nagatsu T.;  
 RT "Molecular cloning of the human Nurrl gene: characterization of the  
 RT human gene and cDNAs.";  
 RL Gene 230:233-239 (1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99234103; PubMed=10216261;  
 RA Torii T., Kawarai T., Nakamura S., Kawakami H.;  
 RT "Organization of the human orphan nuclear receptor Nurrl gene.";  
 RL Gene 230:225-233 (1999).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin;  
 RX MEDLINE=2388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haile F.,  
 RA Diatchenko L., Marusik A., Farmer A.A., Rubin G.M., Hong L.,  
 RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ueidi T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loughran N.A., Peters G.J., Abramson R.D., Mulhally S.J.,  
 RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.T., Skalska U., Smalins D.E.,  
 RA Schnerch A., Schein U.E., Jones S.J.M., Maira M.A.;

RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences";  
 CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- FUNCTION: PROBABLE NUCLEAR RECEPTOR. MAY FUNCTION AS A GENERAL  
 CC COACTIVATOR OF GENE TRANSCRIPTION. DETECTION OF THE PROTEIN IN  
 CC THE BRAIN, INDICATES THAT IT IS NOT MANDATORILY ASSOCIATED WITH  
 CC CELL CYCLE PROGRESSION.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN A NUMBER OF CELL LINES OF T CELL,  
 CC B CELL AND FIBROBLAST ORIGIN. STRONG EXPRESSION IN BRAIN TISSUE.  
 CC -1- DEVELOPMENTAL STAGE: RAPIDLY AND ONLY VERY TRANSIENTLY EXPRESSED  
 CC AFTER CELL ACTIVATION, DURING THE G0-G1 TRANSITION OF THE CELL  
 CC CYCLE.  
 CC -1- SIMILARITY: Belongs to the nuclear hormone receptor family. NR4  
 CC subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X75918; CAA53518.1; -;  
 CC EMBL: AB017586; BAA75666.1; -;  
 CC EMBL: AB019433; BAA77328.1; -;  
 CC EMBL: BC009288; AAO9288.1; -;  
 CC PIR: A57040; A57040.  
 CC HSSP: P19793; 2NLL.  
 CC TRANSFAC: T02742; -;  
 CC Gene: HGNC:7981; NR4A2.  
 CC MIM: 601828; -;  
 CC DR GO: GO:0005634; C:nucleus; TAS.  
 CC DR GO: GO:0004879; F:ligand-dependent nuclear receptor activity; TAS.  
 CC DR GO: GO:0006960; P:antimicrobial humoral response (sensu Inver. . .); TAS.  
 CC DR InterPro: IPR000536; Hormone\_rec\_1lg.  
 CC DR InterPro: IPR001723; Stdhnm\_receptor.  
 CC DR InterPro: IPR008946; Str\_ncl\_receptor.  
 CC DR InterPro: IPR001628; Znf\_C4steroid.  
 CC DR Pfam: PF00104; hormone\_rec; 1.  
 CC DR Pfam: PF00105; zf-C4; 1.  
 CC DR PRINTS: PRO0398; STRDHOMONER.  
 CC DR PRINTS: PRO0047; STROIDFINGER.  
 CC DR PRODOM: PD000035; Znf\_C4steroid; 1.  
 CC DR SMART: SM00430; HOL1; 1.  
 CC DR SMART: SM00399; Znf\_C4; 1.  
 CC DR PROSITE: PS00031; NUCLEAR RECEPTOR; 1.  
 CC DR Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
 CC zinc-finger.  
 CC KW DOMAIN 80 138 GIN-RICH.  
 CC FT DOMAIN 127 233 PRO-RICH.  
 CC FT FT 263 328 NUCLEAR RECEPTOR-TYPE.  
 CC FT ZN\_FING 263 328 C4-TYPE.  
 CC FT ZN\_FING 299 323 C4-TYPE.  
 CC FT DOMAIN 345 391 PRO-RICH.  
 CC FT DOMAIN 409 459 LIGAND-BINDING (POTENTIAL).  
 CC SQ SEQUENCE 598 AA; 66590 MW; 28D8199AC3E211 CRC64;  
 CC  
 CC Query Match 77.3%; Score 34; DB 1; Length 598;  
 CC Best Local Similarity 100.0%; Pred. No. 2; 2e+02;  
 CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Orphan nuclear receptor NURR1 (NUR-related factor 1).  
 GN NR4A2 OR NURR1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10990;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=93149122; PubMed=1491694;  
 RA Law S.W., Connely O.M., Demayo F.J., O'Malley B.W.;  
 RT Identification of a new brain-specific transcription factor, NURR1.;  
 RL Mol. Endocrinol. 6:2129-2135(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SvJ;  
 RX MEDLINE=97288523; PubMed=9143501;  
 RA Castillo S.O., Xiao O., Lyu M.S., Kozak C.A., Nikodem V.M.;  
 RT Organization, sequence, chromosomal localization, and promoter  
 RT identification of the mouse orphan nuclear receptor Nurrl gene.;  
 RL Genomics 41:250-257(1997).  
 CC -1- FUNCTION: PROBABLE NUCLEAR RECEPTOR THAT REGULATES GENE EXPRESSION  
 CC IN BRAIN TISSUE.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=NURR1;  
 CC IsoId=Q06219-1; Sequence=Displayed;  
 CC Name=NURR1A;  
 CC IsoId=Q06219-2; Sequence=VSP\_003710, VSP\_003711;  
 CC -1- TISSUE SPECIFICITY: Brain.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSION BEGINS IN THE EMBRYO, INCREASES IN  
 CC NEONATES AND DECREASES IN THE ADULT.  
 CC -1- SIMILARITY: Belongs to the nuclear hormone receptor family. NR4  
 CC subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: SS3744; AAB25138.1; -;  
 CC EMBL: U86783; AAC53153.1; -;  
 CC EMBL: U86783; AAC53154.1; -;  
 CC PIR: A46225; A46225.  
 CC HSSP: P19793; 2NLL.  
 CC TRANSFAC: T04312; -;  
 CC MGD: MGI:1352456; NR4A2.  
 CC DR GO: GO:0005634; C:nucleus; IDA.  
 CC DR GO: GO:0005634; C:nucleus; TAS.  
 CC DR GO: GO:0007399; P:neurogenesis; IMP.  
 CC DR InterPro: IPR000536; Hormone\_rec\_1lg.  
 CC DR InterPro: IPR001723; Stdhnm\_receptor.  
 CC DR InterPro: IPR008946; Str\_ncl\_receptor.  
 CC DR InterPro: IPR001628; Znf\_C4steroid.  
 CC DR Pfam: PF00104; hormone\_rec; 1.  
 CC DR Pfam: PF00105; zf-C4; 1.  
 CC DR PRINTS: PRO0398; STRDHOMONER.  
 CC DR PRINTS: PRO0047; STROIDFINGER.  
 CC DR PRODOM: PD000035; Znf\_C4steroid; 1.  
 CC DR SMART: SM00430; HOL1; 1.  
 CC DR SMART: SM00399; Znf\_C4; 1.  
 CC DR PROSITE: PS00031; NUCLEAR RECEPTOR; 1.  
 CC DR Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
 CC zinc-finger; Alternative splicing.  
 CC KW DOMAIN 80 138 GIN-RICH.  
 CC FT DOMAIN 127 233 PRO-RICH.  
 CC FT ZN\_FING 263 328 NUCLEAR RECEPTOR-TYPE.  
 CC FT ZN\_FING 263 283 C4-TYPE.

```

FT ZN_FING 299 323 C4-TYPE.
FT DOMAIN 346 391 PRO-RICH.
FT DOMAIN 409 459 LIGAND-BINDING (POTENTIAL).
FT VARSPIC 455 455 S -> I (in isoform NURR1A).
FT VARSPIC 456 598 /FtId=VSP_003710.
FT VARSPIC 456 598 Missing (in isoform NURR1A).
FT VARSPIC 456 598 /FtId=VSP_003711.
SQ SEQUENCE 598 AA; 66592 MW; DIAC5E2FFB8D366 CRC64;

Query Match 77.3%; Score 34; DB 1; Length 598;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPP 6
Db 123 KPSSPP 128

RESULT 44
NR42_RAT
ID NR42_RAT STANDARD; PRT; 598 AA.
AC 007917; 035865;
DT 01-OCT-1994 (Rel. 30, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Orphan nuclear receptor NURR1 (NUR-related factor 1) (Regenerating
DE liver nuclear receptor 1) (RNR-1) (SL-322) (Nuclear orphan receptor
DE HZF-3).
GN NR42 OR NURR1 OR RNR1 OR HZF-3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Hippocampus;
RX MEDLINE=94335560; PubMed=7914660;
RA de Ortiz S., Cannon M.M., Jamieson G.A.;
RT "Expression of nuclear hormone receptors within the rat hippocampus:
RT identification of novel orphan receptors.";
RL Brain Res. Mol. Brain Res. 23:278-283 (1994).
[2]
RN SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93232045; PubMed=8473329;
RA Seearce L.M., Laz T.M., Hazel T.G., Lau L.F., Taub R.;
RT "NUR-1, a nuclear receptor in the NGFI-B/Nur77 family that is rapidly
RT induced in regenerating liver.";
RL J. Biol. Chem. 268:885-886 (1993).
[3]
RN SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=97364974; PubMed=9221923;
RA Xing G.O., Zhang L.X., Zhang L., Heynen T., Li X.L., Smith M.A.,
RA Weiss S.R.B., Feldman A.N., DeCera-Wadleigh S., Chuang D.M.;
RT "Rat nurrl is prominently expressed in perirhinal cortex, and
RT differentially induced in the hippocampal dentate gyrus by
RT electroconvulsive vs. kindled seizures.";
RL Brain Res. Mol. Brain Res. 47:251-261 (1997).
-1- FUNCTION: PROBABLE NUCLEAR RECEPTOR. MAY CONFER LIVER-SPECIFIC
-1- REGULATION OF DELAYED-EARLY GENES INDUCED LATER IN THE G1 PHASE
-1- OF REGENERATION ALONG WITH HMR.
-1- SUBCELLULAR LOCATION: Nuclear.
-1- TISSUE SPECIFICITY: NOT EXPRESSED IN QUIESCENT LIVER BUT IS
-1- RAPIDLY INDUCED FOLLOWING PARTIAL HEPATECTOMY AND IS SPECIFIC TO
-1- HEPATIC GROWTH AS IT IS NOT INDUCED IN OTHER MITOGEN-TREATED
-1- CELLS. EXPRESSED AT VERY LOW LEVELS IN THE LUNG, SPLEEN AND
-1- STOMACH AND AT HIGH LEVELS IN THE BRAIN.
-1- SIMILARITY: Belongs to the nuclear hormone receptor family. NR4
-1- subfamily.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U01146; AAC52143.1; -
DR EMBL; L08595; AAA42058.1; -
DR EMBL; U72345; AAC53315.1; -
DR PIR; I84692; I84692.
DR HSSP; P19793; ZML.
DR TRANSFAC; T04743; -
DR InterPro; IPR000536; Hormone_rec_1g.
DR InterPro; IPR001723; Steroid_receptor.
DR InterPro; IPR008946; Str_ncl_receptor.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec_1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STEROIDFINGER.
DR Prodom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOL1.1.
DR SMART; SM00399; ZNF_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR Receptor; Transcription regulation; DNA-binding; Nuclear protein;
DR Zinc-finger.
KM
FT DOMAIN 80 138 GLN-RICH.
FT DOMAIN 127 233 PRO-RICH.
FT DNA_BIND 263 328 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 263 283 C4-TYPE.
FT ZN_FING 299 323 C4-TYPE.
FT DOMAIN 346 391 PRO-RICH.
FT DOMAIN 409 459 PRO-RICH.
FT CONFLICT 182 182 LIGAND-BINDING (POTENTIAL).
FT CONFLICT 182 182 R -> P (IN REF. 2).
FT CONFLICT 250 250 S -> P (IN REF. 3).
FT CONFLICT 252 252 S -> A (IN REF. 1).
FT CONFLICT 362 362 V -> G (IN REF. 2).
FT CONFLICT 364 371 LISALVRA -> DOCPROT (IN REF. 2).
FT CONFLICT 434 434 D -> S (IN REF. 2).
SQ SEQUENCE 598 AA; 66621 MW; 0A54ED59FB70DEA8 CRC64;

Query Match 77.3%; Score 34; DB 1; Length 598;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPP 6
Db 123 KPSSPP 128

RESULT 45
Y013_SCHPO
ID Y013_SCHPO STANDARD; PRT; 615 AA.
AC 013681;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C737.03c in chromosome III.
GN SPC737.03C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxId=4896;
RN
RN SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sguros J., Peat N., Hayles J., Baker S., Baahman D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,

```



RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Genies S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holtroyd S., Honesty T., Howarth S., Huckle E.J., Hunt S., Jagals K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., O'Neill C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabbittowitch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmons M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodard J., Wolcraet G., Aert R., Rooben J., Grympey B.,  
RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabell C., Fuchs M., Fritsch C., Holzer B., Moesti D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Medler H., Manburt R., Purnelle B.,  
RA Coffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Gallbert F., Aves S.U., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez S., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Foreburg S.L.,  
RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,  
RA "The genome sequence of *Schizosaccharomyces pombe*."  
RL Nature 415:871-880(2002).  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AL031546; CAA20860.1; -.  
CC DR PIR: T41576; T41576.  
CC DR GeneDB: SPombe; SPCC737.03c; -.  
CC KW Hypothetical protein; Transmembrane.  
CC FT TRANSMEM 192 212 POTENTIAL.  
CC FT TRANSMEM 247 267 POTENTIAL.  
CC FT TRANSMEM 323 343 POTENTIAL.  
CC FT TRANSMEM 563 583 POTENTIAL.  
CC FT TRANSMEM 586 606 POTENTIAL.  
CC FT DOMAIN 88 91 POLY-SER.  
CC SQ SEQUENCE 615 AA; 70433 MW; DSCA8589441064E CRC64;  
  
CC Query Match 77.3%; Score 34; DB 1; Length 615;  
CC Best Local Similarity 85.7%; Pred. No. 2.3e+02;  
CC Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
CC QY 1 KPSPPE 7  
CC DB 344 KPSPPE 350  
  
CC RESULT 46  
CC ID CN4A\_HUMAN STANDARD; PRT; 886 AA.  
CC AC P27815; O75522; O76092; Q16255; Q16691; Q8WU03;  
CC DT 01-AUG-1992 (Rel. 23, Created)  
CC DT 01-OCT-1996 (Rel. 34, Last sequence update)  
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)  
CC DE CAMP-specific 3',5'-cyclic phosphodiesterase 4A (EC 3.1.4.17)  
CC DE (DPDB2) (PDB46).  
CC GN PD34A.  
CC OS Homo sapiens (Human).  
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
CC OX NCBI\_TaxID=9606;  
CC RN [1]  
CC RP SEQUENCE FROM N.A. (ISOFORM 1).  
CC RX MEDLINE=94019330; PubMed=8413254;  
CC RA Bolger G., Michael T., Martins T., St John T., Steiner B.,  
CC Rodgers L., Riggs M., Wigler M., Ferguson K.;

RT "A family of human phosphodiesterases homologous to the dunce  
RT learning and memory gene product of *Drosophila melanogaster* are  
RT potential targets for antidepressant drugs."  
RL Mol. Cell. Biol. 13:6558-6571(1993).  
  
CC [2]  
CC RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5).  
CC RX TISSUE=Brain;  
CC MEDLINE=98333959; PubMed=9677330;  
CC RA Sullivan M., Rana G., Begg F., Gordon L., Olsen A.S., Houslay M.D.;  
CC RT "Identification and characterization of the human homologue of the  
CC short PD34A cAMP-specific phosphodiesterase RDI (PD34A) by analysis  
CC of the human HSP60A gene locus located at chromosome 19p13.2.";  
CC RL Biochem. J. 333:693-703(1998).  
  
CC [3]  
CC RP SEQUENCE OF 112-886 FROM N.A.  
CC RC TISSUE=Monocytes;  
CC RX MEDLINE=90258854; PubMed=2160582;  
CC RA Livi G.P., Kmetz P., McHale M.M., Cieplinski L.B., Sathe G.M.,  
CC RA Taylor D.P., Davis R.L., Torphy T.J., Balcarak J.M.;  
CC RT "Cloning and expression of cDNA for a human low-Km, rolipram-sensitive  
CC cyclic AMP phosphodiesterase.";  
CC RL Mol. Cell. Biol. 10:2678-2686(1990).  
  
CC [4]  
CC RP REVISIONS TO REF. 3.  
CC RA McLaughlin M.M.;  
CC RL Submitted (JAN-1991) to the EMBL/GenBank/DBJ databases.  
  
CC [5]  
CC RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 5).  
CC RX MEDLINE=95290008; PubMed=7772058;  
CC RA Horton Y.M., Sullivan M., Houslay M.D.;  
CC RT "Molecular cloning of a novel splice variant of human type IVA  
CC (PDE-IVA) cyclic AMP phosphodiesterase and localization of the gene  
CC to the p13.2-q12 region of human chromosome 19.";  
CC RL Biochem. J. 308:683-691(1995).  
  
CC [6]  
CC RP SEQUENCE FROM N.A. (ISOFORM 3).  
CC RX MEDLINE=95194817; PubMed=7888306;  
CC RA Sullivan M., Egerton M., Shakur Y., Margardsen A., Houslay M.D.;  
CC RT "Molecular cloning and expression, in both COS-1 cells and S.  
CC cerevisiae, of a human cytosolic type-IVA, cyclic AMP specific  
CC phosphodiesterase (hPDE-IVA-h6.1).";  
CC RL Cell. Signal. 6:793-812(1994).  
  
CC [7]  
CC RP SEQUENCE FROM N.A. (ISOFORM 4).  
CC RC TISSUE=Brain;  
CC RX MEDLINE=2388257; PubMed=12477932;  
CC RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
CC RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
CC RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
CC RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
CC RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
CC RA Stapleton W., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,  
CC RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
CC RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Millaly S.J.,  
CC RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
CC RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
CC RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
CC RA Whiting M., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
CC RA Blakesley A.C., Touchman J.W., Green E.D., Dickinson M.C.,  
CC RA Butterfield Y.S.N., Krzywicki M.I., Skalek U., Smalins D.E.,  
CC RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
CC RT "Generation and initial analysis of more than 15,000 full-length  
CC human and mouse cDNA sequences.";  
CC RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
  
CC -1- CATALYTIC ACTIVITY: Adenosine 3',5'-cyclic phosphate + H(2)O =  
CC adenosine 5'-phosphate.  
CC -1- ENZYME REGULATION: Inhibited by rolipram.  
CC -1- PATHWAY: Cyclic nucleotide metabolism.  
CC -1- SUBCELLULAR LOCATION: Isoform 4 has propensity for association  
CC with membranes.  
CC -1- ALTERNATIVE PRODUCTS:



CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL, 272841; CA97057.1; --.

DR PIR, S643501; S64350.

DR HSSP, Q92831; 1B91.

DR GenOnline, 141368; --.

DR SGD, S0003288; RSC1.

DR GO, GO:0005679; C:nucleosome remodeling complex; IPL.

DR GO, GO:0006338; P:chromatin modeling; IDA.

DR InterPro, IPR001025; BAH.

DR InterPro, IPR001487; Bromodomain.

DR Pfam, PF01426; BAH; 1.

DR Pfam, PF00439; bromodomai; 2.

DR PRINTS, PRO0503; BROMODOMAIN.

DR SMART, SM00439; BAH; 1.

DR SMART, SM00297; BROMO; 2.

DR PROSITE, PS00633; BROMODOMAIN 1; 1.

DR PROSITE, PS00633; BROMODOMAIN 2; 2.

KW Chromatin regulator; Nuclear Protein; Bromodomai; Repeat.

FT DOMAIN 27 95 BROMODOMAIN 1.

FT DOMAIN 235 325 BROMODOMAIN 2.

SQ SEQUENCE 928 AA; 106669 MW; EFP80922FC08EC27 CRC64;

Query Match 77.3%; Score 34; DB 1; Length 928;  
Best local Similarity 62.5%; Pred. No. 3.6e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KPSSPEE 8  
:|:|||||  
Db 575 RPMDPEE 582

RESULT 48

G2D1 HUMAN STANDARD; PRT; 959 AA.

ID Q9UHL9; O95444; O9UHK8; Q9U191;

AC Q9UHL9; O95444; O9UHK8; Q9U191;

DT 28-FEB-2003 (Rel. 41; Created)

DT 28-FEB-2003 (Rel. 41; Last sequence update)

DT 10-OCT-2003 (Rel. 42; Last annotation update)

DE General transcription factor II-I repeat domain-containing protein 1

DE (GTF2I repeat domain containing protein 1) (Muscle TFIID repeat domain-containing protein 1) (General transcription factor IIT) (Slow-muscle-fiber enhancer binding protein) (USE B1 binding protein)

DE (MusTRD/BER) (Williams-Beuren syndrome chromosome region 11 protein).

GN GTF2IRD1 OR MUSTRD1 OR GTF3 OR CREAM1 OR RBAP2 OR WBSCR11 OR WBSCR12.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxId=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM 2).

RC TISSUE=Muscle;

RX MEDLINE=98449952; PubMed=9774679;

RA O'Mahoney J.V., Guven K.L., Lin J., Joya J.E., Robinson C.S., Wade R.P., Hardeman E.C.;

RT "Identification of a novel slow-muscle-fiber enhancer binding protein, MUSTRD1.";

RT Mol. Cell. Biol. 18:6641-6652(1998).

RL [2]

RP SEQUENCE FROM N.A. (ISOFORM 2).

RC TISSUE=Colon carcinoma;

RX MEDLINE=99216421; PubMed=10198167;

RA Osborne L.R., Campbell T., Daradich A., Scherer S.W., Tsui L.-C.;

RT "Identification of a putative transcription factor gene (WBSCR11) that is commonly deleted in Williams-Beuren syndrome.";

RL Genomics 57:279-284(1999).

RN [3]

RP SEQUENCE FROM N.A. (ISOFORM 1).

RC TISSUE=Fetal brain;

RX MEDLINE=20037629; PubMed=10573005;

RA Tasabehji M., Carette M., Wilmot C., Donnai D., Read A.P., Metcalfe K.;

RT "A transcription factor involved in skeletal muscle gene expression is deleted in patients with Williams syndrome.";

RT Eur. J. Hum. Genet. 7:737-747(1999).

RL [4]

RP SEQUENCE FROM N.A. (ISOFORM 2).

RX MEDLINE=20044629; PubMed=10575229;

RA Franke Y., Peoples R.J., Francke U.;

RT "Identification of GTF2IRD1, a putative transcription factor within the Williams-Beuren syndrome deletion at 7q11.23.";

RL Cytogenet. Cell Genet. 86:236-304(1999).

RN [5]

RP SEQUENCE FROM N.A. (ISOFORM 1), INTERACTION WITH RB1, AND MUTAGENESIS.

RC TISSUE=Cervical carcinoma, Fetal spleen, and Placenta;

RX MEDLINE=20115113; PubMed=10642537;

RA Yan X., Zhao X., Qian M., Guo N., Gong X., Zhu X.;

RT "Characterization and gene structure of a novel retinoblastoma-protein-associated protein similar to the transcription regulator TFIID-1.";

RL Biochem. J. 345:749-757(2000).

RL [6]

RP SEQUENCE FROM N.A. (ISOFORM 1).

RC TISSUE=Dermis;

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ustin T.B., Toshitoki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaney S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield A.S.N., Krzywicki M.I., Skalska U., Smalins D.E., Scheraga J., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RL [7]

RP FUNCTION.

RX MEDLINE=21332325; PubMed=11438732;

RA Tussie-Luna M.I., Bayarashan D., Ruddle F.H., Roy A.L.;

RT "Repression of TFIID-dependent transcription by nuclear exclusion.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:7789-7794(2001).

CC -1- FUNCTION: May be a transcription regulator involved in cell-cycle progression and skeletal muscle differentiation. May repress GTF2I transcriptional functions, by preventing its nuclear residency, or by inhibiting its transcriptional activation. May contribute to slow-twitch fiber type specificity during myogenesis and in regenerating muscles. Binds tropomyosin I slow-muscle fiber enhancer (USE B1). Binds specifically and with high affinity to the EFG sequences derived from the early enhancer of HOXB8 (By similarity).

CC -1- SUBUNIT: Interacts with the retinoblastoma protein (RB1) via its C-terminus.

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=2;

CC Name=1;

CC IsoId=Q9UHL9-1; Sequence=Displayed;

CC Name=2;

CC IsoId=Q9UHL9-2; Sequence=VSP\_003873;

CC -1- TISSUE SPECIFICITY: Highly expressed in adult skeletal muscle, heart, fibroblast, bone and fetal tissues. Expressed at lower levels in all other tissues tested.

CC -1- DEVELOPMENTAL STAGE: Highly expressed in developing and regenerating muscles, at the time of myofiber diversification.

CC -1- DOMAIN: The N-terminal half may have an activating activity.

```

CC -1- DISEASE: Haploinsufficiency of GTF2IRD1 may be the cause of
CC certain cardiovascular and musculo-skeletal abnormalities observed
CC in Williams-Beuren syndrome (WBS), a rare developmental disorder.
CC It is a contiguous gene deletion syndrome involving genes from
CC chromosome band 7q11.23
CC -1- SIMILARITY: BELONGS TO THE TFIID-I FAMILY.
CC -1- SIMILARITY: Contains 5 GTF2I-like repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF18270; AAD14687.2; -
DR EMBL: AF104923; AAD27668.1; -
DR EMBL: AF151354; AAF19786.1; -
DR EMBL: AF156489; AAF17358.1; -
DR EMBL: AF089107; AAF21796.1; -
DR EMBL: BC018136; AAF18136.1; -
DR Genew: HGNC:4661; GTF2IRD1.
DR MIM: 604318; -
DR MIM: 194050; -
DR GO: GO:0005634; C:nucleus; NAS.
DR GO: GO:0003705; F:RNA polymerase II transcription factor acti...; NAS.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR InterPro: IPR004212; GTF2I.
DR Pfam: PF02946; GTF2I; 5.
DR Transcription regulation; Developmental protein; DNA-binding;
DR Nuclear protein; Repeat; Alternative splicing; Polymorphism;
DR Williams-Beuren syndrome.
DR REPEAT 128 203 GTF2I 1.
DR REPEAT 351 426 GTF2I 2.
DR REPEAT 565 640 GTF2I 3.
DR REPEAT 705 780 GTF2I 4.
DR REPEAT 802 877 GTF2I 5.
DR DOMAIN 898 905 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
DR DOMAIN 906 930 SER-RICH.
DR VARSPLIC 656 670 Missing (in isoform 2).
DR VARIANT 652 652 /FTId=VSP_003873;
DR M -> V (in dbSNP:2301895).
DR MUTAGEN 898 959 /FTId=VAR_013446.
DR CONFLICT 111 MISSING: CYTOPLASMIC LOCALIZATION.
DR CONFLICT 378 G -> S (IN REF. 1 AND 2).
DR CONFLICT 378 R -> Q (IN REF. 5).
DR SEQUENCE 959 AA; 106057 MW; 7DA3097879701540 CRC64;
SQ
Query Match 77.3%; Score 34; DB 1; Length 959;
Best Local Similarity 62.5%; Pred. No. 3.7e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 KPSSPPE 8
Db 445 EPASPED 452

```

```

RESULT 49
MRTB_MOUSE STANDARD; PRT; 1080 AA.
AC P59759;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myocardin-related transcription factor B (MRTF-B).
GN MRTF-B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A., CHARACTERIZATION, FUNCTION, TISSUE SPECIFICITY,

```

```

RP DEVELOPMENTAL STAGE, AND INTERACTION WITH SRF.
RC STRAIN=C57BL/6;
RX MEDLINE=22317395; PubMed=12397177;
RA Wang D.-Z., Li S., Hockeimer D., Sutherland L., Wang Z., Schratz G.,
RT Richardson J.A., Nordheim A., Olson E.N.;
RT "Potentiation of serum response factor activity by a family of
RT myocardin-related transcription factors."
RL Proc. Natl. Acad. Sci. U.S.A. 99:14855-14860(2002).
CC -1- FUNCTION: Poor transcriptional factor which uses the canonical
CC single or multiple CARG boxes DNA sequence. Acts as a cofactor of
CC serum response factor (SRF) with the potential to modulate SRF
CC target genes.
CC -1- SUBUNIT: Interacts with SRF.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- TISSUE SPECIFICITY: Widely expressed. High expression in heart,
CC brain and testis. Lower expression in lung, liver and kidney.
CC -1- DEVELOPMENTAL STAGE: Detected throughout the embryo at 10.5 dpc.
CC High expression in epithelial cells of the lung, kidney, bladder,
CC colon, testis, in the smooth muscle of the colon and small
CC intestines, and in the mesenchymal cells adjacent to the olfactory
CC epithelium at 15.5 dpc.
CC -1- DOMAIN: The N-terminal region is required for nuclear localization
CC and the C-terminal region mediates transcriptional activity (By
CC similarity). Contains 3 RPEL repeats.
CC -1- SIMILARITY: Contains 1 SAP domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF332598; AAN33042.1; -
DR InterPro: IPR004018; RPEL_repeat.
DR InterPro: IPR003034; SAP.
DR Pfam: PF02755; RPEL; 3.
DR Pfam: PF02037; SAP; 1.
DR SMART: SM00707; RPEL; 3.
DR SMART: SM00513; SAP; 1.
DR PROSITE: PS00800; SAP; 1.
DR Transcription regulation; Nuclear protein; Coiled coil; Repeat.
DR REPEAT 40 65 RPEL 1.
DR REPEAT 84 109 RPEL 2.
DR REPEAT 128 153 RPEL 3.
DR DOMAIN 320 346 GLN-RICH.
DR DOMAIN 383 417 SAP.
DR DOMAIN 539 594 COILED COIL (POTENTIAL).
DR DOMAIN 665 806 GLN-RICH.
DR SEQUENCE 1080 AA; 117546 MW; 1781DBEF34517DAC CRC64;
SQ
Query Match 77.3%; Score 34; DB 1; Length 1080;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KPSSP 6
Db 836 KPSSP 841

```

```

RESULT 50
MRTB_HUMAN STANDARD; PRT; 1088 AA.
AC Q9ULH7; O86MW2; Q8N226;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myocardin-related transcription factor B (MRTF-B).
GN MRTF-B OR KIAA1243.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CC NCBI\_TaxID=9606;  
 CC [1] TaxID=9606;  
 CC SEQUENCE FROM N.A. (ISOFORM 3).  
 CC TISSUE=Thymus;  
 CC Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,  
 CC Arima M., Musashino K., Yuki H., Hara H., Sugiyama T., Irie R.,  
 CC Otsuki T., Sato H., Ota T., Makamatsu A., Ishii S., Yamamoto J.,  
 CC Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,  
 CC Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,  
 CC Wagaetsuma M., Murakawa K., Kanemori K., Sugiyama A., Kawakami B.,  
 CC Suzuki Y., Sugano S., Nagahara K., Masuho Y., Nagai K., Isogai T.,  
 CC "NEO human cDNA sequencing project."  
 CC Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 CC [2]  
 CC SEQUENCE FROM N.A. (ISOFORM 2).  
 CC TISSUE=Medulla oblongata;  
 CC MEDLINE=2238257; PubMed=12477932;  
 CC Klausner R.D., Collins F.S., Wagner L., Spemmen C.M., Schuler G.D.,  
 CC Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 CC Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 CC Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 CC Diatchenko L., Matulis A., Farmer A.A., Rubin G.M., Hong L.,  
 CC Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 CC Brownstein M.J., Udell T.B., Tohyuki S., Carninci P., Prange C.,  
 CC Raha S.S., Lonnellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 CC Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunatane P.H.,  
 CC Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 CC Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 CC Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 CC Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 CC Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 CC Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 CC Buterfield Y.S.N., Krzywicki M.I., Skalek U., Smilans D.E.,  
 CC Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 CC "Generation and initial analysis of more than 15,000 full-length human  
 CC and mouse cDNA sequences."  
 CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC [3]  
 CC SEQUENCE OF 261-1088 FROM N.A. (ISOFORM 1).  
 CC TISSUE=Brain;  
 CC MEDLINE=20039619; PubMed=10574462;  
 CC Nagase T., Ishikawa K.-I., Kikuno R., Hirosewa M., Nomura N.,  
 CC Ohara O.,  
 CC "Prediction of the coding sequences of unidentified human genes. XV.  
 CC The complete sequences of 100 new cDNA clones from brain which code  
 CC for large proteins in vitro."  
 CC DNA Res. 6:337-345(1999).  
 CC -1- FUNCTION: Poor transcriptional factor which uses the canonical  
 CC single or multiple CARG boxes DNA sequence. Acts as a cofactor of  
 CC serum response factor (SRF) with the potential to modulate SRF  
 CC target genes (By similarity).  
 CC -1- SUBUNIT: Interacts with SRF (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Comment=Full isoforms so far detected are isoform 2 and isoform  
 CC 3;  
 CC Name=1;  
 CC IsoId=Q9ULH7-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q9ULH7-2; Sequence=VSP\_007653, VSP\_007654, VSP\_007655;  
 CC Note=No experimental confirmation available;  
 CC Name=3;  
 CC IsoId=Q9ULH7-3; Sequence=VSP\_007656, VSP\_007657;  
 CC Note=No experimental confirmation available;  
 CC -1- DOMAIN: The N-terminal region is required for nuclear localization  
 CC and the C-terminal region mediates transcriptional activity (By  
 CC similarity).  
 CC -1- SIMILARITY: Contains 3 RPEL repeats.  
 CC -1- SIMILARITY: Contains 1 SAP domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AK093577; BAC04200.1; -  
 CC EMBL: BC047761; AA047761.1; -  
 CC EMBL: AB033069; BA06557.2; -  
 CC InterPro: IPR004018; RPEL repeat.  
 CC InterPro: IPR003034; SAP.  
 CC Pfam: PF02755; RPEL; 3.  
 CC Pfam: PF02037; SAP; 1.  
 CC SMART: SM00707; RPEL; 3.  
 CC SMART: SM00513; SAP; 1.  
 CC PROSITE: PS00800; SAP; 1.  
 CC Transcription regulation; Nuclear protein; Coiled coil; Repeat;  
 CC Alternative splicing.  
 CC REPEAT 40 65  
 CC REPEAT 84 109  
 CC REPEAT 128 153  
 CC DOMAIN 389 423  
 CC DOMAIN 545 601  
 CC DOMAIN 671 787  
 CC VARSPLIC 1 41  
 CC --> M (in isoform 2).  
 CC /FTId=VSP\_007653.  
 CC PLNDKNSNGSNALNATPNTPTPNTSTPVRKGPPLSSID  
 CC DLKVSLEKTELKRGHPVSGTKEDLIERLKPYOVNNSGIA  
 CC MYRVACIPROFLSKIGSEFLQVRNAPSOLFICILLEHQ  
 CC NSTRCEKSVSSIIPIGINS (in isoform 2).  
 CC /FTId=VSP\_007654.  
 CC Missing (in isoform 2).  
 CC /FTId=VSP\_007655.  
 CC PLNDKNSNGSNALNATPNTPTPNTSTPVRKGPPLSSID  
 CC DLKVSLEKTELKRGHPVSGTKEDLIERLKPYOVNNSGIA  
 CC AGGIVAVSSAIVTSNPEVTALPTTINH -> YGGAHVI  
 CC LNRGSPVRRNRYKLEKVECHFLFVSNDFPFIYAHYV  
 CC SEVHMYRVACIPROFLSKIGSEFLQVRNAPSOLFICIL  
 CC LEHONSTRCEKSVSSIIPIGINS (in isoform 3).  
 CC /FTId=VSP\_007656.  
 CC Missing (in isoform 3).  
 CC /FTId=VSP\_007657.  
 CC FT CONFLICT 266 266 K -> R (in RRF. 1).  
 CC SQ SEQUENCE 1088 AA; 118126 MW; 0CA4A52A115C0C83 CRC64;  
 CC Query Match 77.3%; Score 34; DB 1; Length 1088;  
 CC Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
 CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC QY 1 KPSSPP 6  
 CC Db 844 KPSSPP 849

Search completed: June 18, 2004, 12:58:58  
 Job time : 16 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model1

Run on: June 18, 2004, 12:55:54 ; Search time 39 Seconds  
(without alignments)  
64.722 Million cell updates/sec

Title: US-10-655-201-2

Perfect score: 44  
Sequence: 1 KPSPPEB 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 500 summaries

Database :

SPTREMBL\_25: \*  
1: sp archaea: \*  
2: sp bacteria: \*  
3: sp fungi: \*  
4: sp human: \*  
5: sp invertebrate: \*  
6: sp mammal: \*  
7: sp mhc: \*  
8: sp organelle: \*  
9: sp phage: \*  
10: sp plant: \*  
11: sp rodent: \*  
12: sp virus: \*  
13: sp vertebrate: \*  
14: sp unclassified: \*  
15: sp\_rv1rus: \*  
16: sp\_bacteriap: \*  
17: sp\_archaeap: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	93.2	566	10 Q8H611	Q8H611 zea mays (m
2	39	88.6	390	10 Q8R289	Q8R289 oryza sativ
3	39	88.6	802	10 Q8L775	Q8L775 arabidopsis
4	39	88.6	990	10 Q9LN08	Q9LN08 arabidopsis
5	38	86.4	87	15 Q38534	Q38534 human immun
6	38	86.4	215	11 Q9J003	Q9J003 mus musculu
7	38	86.4	216	4 Q8N5C9	Q8N5C9 homo sapien
8	38	86.4	460	11 Q8R4Z2	Q8R4Z2 mus musculu
9	38	86.4	460	11 Q8K4H6	Q8K4H6 mus musculu
10	38	86.4	461	4 Q96NC3	Q96NC3 mus musculu
11	38	86.4	461	4 Q96NC3	Q96NC3 homo sapien
12	38	86.4	607	10 Q9J079	Q9J079 arabidopsis
13	38	86.4	1351	3 Q872K5	Q872K5 neurospora
14	37	84.1	52	15 Q9J3Q8	Q9J3Q8 human immun
15	37	84.1	52	15 Q9J3S0	Q9J3S0 human immun
16	37	84.1	67	15 Q9JF19	Q9JF19 human immun

17	37	84.1	68	15 Q9DYK6	Q9DYK6 human immun
18	37	84.1	73	15 Q9JF25	Q9JF25 human immun
19	37	84.1	68	15 Q38224	Q38224 human immun
20	37	84.1	75	15 Q38232	Q38232 human immun
21	37	84.1	76	15 Q90QC3	Q90QC3 human immun
22	37	84.1	91	15 Q90QH2	Q90QH2 human immun
23	37	84.1	94	15 Q9DYQ4	Q9DYQ4 human immun
24	37	84.1	94	15 Q9DYQ1	Q9DYQ1 human immun
25	37	84.1	94	15 Q9DYQ7	Q9DYQ7 human immun
26	37	84.1	95	15 Q38202	Q38202 human immun
27	37	84.1	95	15 Q38182	Q38182 human immun
28	37	84.1	96	15 Q38195	Q38195 human immun
29	37	84.1	96	15 Q38214	Q38214 human immun
30	37	84.1	96	15 Q38215	Q38215 human immun
31	37	84.1	96	15 Q38208	Q38208 human immun
32	37	84.1	96	15 Q38218	Q38218 human immun
33	37	84.1	96	15 Q38210	Q38210 human immun
34	37	84.1	96	15 Q38193	Q38193 human immun
35	37	84.1	96	15 Q38183	Q38183 human immun
36	37	84.1	96	15 Q38204	Q38204 human immun
37	37	84.1	96	15 Q38221	Q38221 human immun
38	37	84.1	96	15 Q38205	Q38205 human immun
39	37	84.1	96	15 Q38206	Q38206 human immun
40	37	84.1	96	15 Q38198	Q38198 human immun
41	37	84.1	96	15 Q38207	Q38207 human immun
42	37	84.1	96	15 Q38197	Q38197 human immun
43	37	84.1	96	15 Q38184	Q38184 human immun
44	37	84.1	96	15 Q38194	Q38194 human immun
45	37	84.1	96	15 Q38200	Q38200 human immun
46	37	84.1	96	15 Q38211	Q38211 human immun
47	37	84.1	96	15 Q38216	Q38216 human immun
48	37	84.1	96	15 Q38217	Q38217 human immun
49	37	84.1	96	15 Q38213	Q38213 human immun
50	37	84.1	96	15 Q38222	Q38222 human immun
51	37	84.1	96	15 Q38180	Q38180 human immun
52	37	84.1	96	15 Q38189	Q38189 human immun
53	37	84.1	96	15 Q38220	Q38220 human immun
54	37	84.1	96	15 Q38217	Q38217 human immun
55	37	84.1	96	15 Q38192	Q38192 human immun
56	37	84.1	96	15 Q38212	Q38212 human immun
57	37	84.1	96	15 Q38190	Q38190 human immun
58	37	84.1	96	15 Q38209	Q38209 human immun
59	37	84.1	96	15 Q38196	Q38196 human immun
60	37	84.1	98	15 Q38227	Q38227 human immun
61	37	84.1	98	15 Q38234	Q38234 human immun
62	37	84.1	98	15 Q38231	Q38231 human immun
63	37	84.1	110	15 Q38240	Q38240 human immun
64	37	84.1	110	15 Q38243	Q38243 human immun
65	37	84.1	111	15 Q38241	Q38241 human immun
66	37	84.1	111	15 Q38235	Q38235 human immun
67	37	84.1	113	15 Q75695	Q75695 human immun
68	37	84.1	113	15 Q38244	Q38244 human immun
69	37	84.1	113	15 Q38238	Q38238 human immun
70	37	84.1	113	15 Q38242	Q38242 human immun
71	37	84.1	113	15 Q38236	Q38236 human immun
72	37	84.1	420	17 Q8RYP8	Q8RYP8 mechanopyru
73	37	84.1	493	15 Q9J4M7	Q9J4M7 human immun
74	37	84.1	495	15 Q90157	Q90157 human immun
75	37	84.1	495	15 Q7J210	Q7J210 human immun
76	37	84.1	498	15 Q7J213	Q7J213 human immun
77	37	84.1	500	15 Q9Z897	Q9Z897 human immun
78	37	84.1	500	15 Q72626	Q72626 human immun
79	37	84.1	500	15 Q9J0H2	Q9J0H2 human immun
80	37	84.1	500	15 Q72634	Q72634 human immun
81	37	84.1	500	15 Q72627	Q72627 human immun
82	37	84.1	500	15 Q72633	Q72633 human immun
83	37	84.1	500	15 Q72629	Q72629 human immun
84	37	84.1	500	15 Q72635	Q72635 human immun
85	37	84.1	500	15 Q8AUI5	Q8AUI5 human immun
86	37	84.1	500	15 Q727H3	Q727H3 human immun
87	37	84.1	502	15 Q8AOW4	Q8AOW4 human immun
88	37	84.1	515	15 Q9Z900	Q9Z900 human immun
89	37	84.1	517	15 Q71261	Q71261 human immun

90	37	84.1	517	15	Q73329	Q73329 human immun	163	35	79.5	1896	13	Q91AJ1	Q91AJ1 xenopus lae
91	37	84.1	562	5	Q8T052	Q8T052 dirosophila	164	35	79.5	1904	11	Q64699	Q64699 mus musculus
92	37	84.1	843	5	Q8X0M6	Q8X0M6 neurospora	165	35	79.5	2137	4	Q15021	Q15021 homo sapien
93	37	84.1	846	5	Q8MRP6	Q8MRP6 dirosophila	166	34	77.3	17	4	Q9URJ5	Q9URJ5 homo sapien
94	37	84.1	1099	5	Q9VMS5	Q9VMS5 dirosophila	167	34	77.3	52	15	Q71399	Q71399 human immun
95	37	84.1	1109	5	Q9VAX4	Q9VAX4 dirosophila	168	34	77.3	56	15	Q7SPR4	Q7SPR4 human immun
96	37	84.1	1150	5	Q8IMM6	Q8IMM6 dirosophila	169	34	77.3	65	15	Q9DYK5	Q9DYK5 human immun
97	36	81.8	124	16	Q9A473	Q9A473 caulobacter	170	34	77.3	65	15	Q91F09	Q91F09 human immun
98	36	81.8	244	13	Q8UJH7	Q8UJH7 gallus gall	171	34	77.3	66	15	Q38530	Q38530 human immun
99	36	81.8	261	16	Q9Z5K4	Q9Z5K4 mycobacteri	172	34	77.3	66	15	Q38432	Q38432 human immun
100	36	81.8	263	5	Q9VDJ7	Q9VDJ7 dirosophila	173	34	77.3	66	15	Q8AEV2	Q8AEV2 human immun
101	36	81.8	268	10	Q8S2A7	Q8S2A7 oryza sativ	174	34	77.3	66	16	Q8YKD9	Q8YKD9 arabidopsis
102	36	81.8	275	13	Q7SXJ7	Q7SXJ7 brachydantio	175	34	77.3	67	15	Q38431	Q38431 human immun
103	36	81.8	362	4	Q13236	Q13236 homo sapien	176	34	77.3	67	15	Q9DYK3	Q9DYK3 human immun
104	36	81.8	442	10	Q7XN43	Q7XN43 oryza sativ	177	34	77.3	67	15	Q38295	Q38295 human immun
105	36	81.8	502	15	Q7ZJH8	Q7ZJH8 human immun	178	34	77.3	67	15	Q38429	Q38429 human immun
106	36	81.8	519	15	Q7ZJH8	Q7ZJH8 human immun	179	34	77.3	67	15	Q8AEW0	Q8AEW0 human immun
107	36	81.8	574	4	Q9NVB3	Q9NVB3 homo sapien	180	34	77.3	68	15	Q9DYK0	Q9DYK0 human immun
108	36	81.8	622	13	Q7ZWV5	Q7ZWV5 xenopus lae	181	34	77.3	68	15	Q38430	Q38430 human immun
109	36	81.8	655	11	Q9D2Z9	Q9D2Z9 mus musculu	182	34	77.3	68	15	Q9DYK4	Q9DYK4 human immun
110	36	81.8	671	10	Q9FNA9	Q9FNA9 arabidopsis	183	34	77.3	68	15	Q9DYK1	Q9DYK1 human immun
111	36	81.8	698	16	Q83BB6	Q83BB6 coxiella bu	184	34	77.3	68	15	Q9DYK7	Q9DYK7 human immun
112	36	81.8	724	11	Q8BZM6	Q8BZM6 mus musculu	185	34	77.3	69	15	Q9DYK2	Q9DYK2 human immun
113	36	81.8	816	16	Q8XT29	Q8XT29 ralestona, s	186	34	77.3	71	15	Q91F21	Q91F21 human immun
114	36	81.8	851	11	Q9CS19	Q9CS19 mus musculu	187	34	77.3	73	15	Q38186	Q38186 human immun
115	36	81.8	877	11	Q8K332	Q8K332 mus musculu	188	34	77.3	73	15	Q38428	Q38428 human immun
116	36	81.8	1154	4	Q9P2P0	Q9P2P0 homo sapien	189	34	77.3	73	15	Q38188	Q38188 human immun
117	36	81.8	1185	4	Q9NTD2	Q9NTD2 homo sapien	190	34	77.3	73	15	Q38187	Q38187 human immun
118	36	81.8	1329	4	Q86VY3	Q86VY3 homo sapien	191	34	77.3	73	15	Q91F07	Q91F07 human immun
119	36	81.8	1430	11	Q8VHK2	Q8VHK2 rattus norv	192	34	77.3	73	15	Q38185	Q38185 human immun
120	36	81.8	1431	4	Q8WAD9	Q8WAD9 homo sapien	193	34	77.3	74	15	Q9DYI2	Q9DYI2 human immun
121	36	81.8	1598	4	Q9P214	Q9P214 homo sapien	194	34	77.3	74	15	Q90QC1	Q90QC1 human immun
122	36	81.8	1633	11	Q8CFW7	Q8CFW7 mus musculu	195	34	77.3	75	15	Q38523	Q38523 human immun
123	35	79.5	90	16	Q91SC1	Q91SC1 pseudomons	196	34	77.3	75	15	Q38438	Q38438 human immun
124	35	79.5	120	4	Q95608	Q95608 homo sapien	197	34	77.3	75	15	Q38491	Q38491 human immun
125	35	79.5	189	11	Q8CG36	Q8CG36 mus sp. fib	198	34	77.3	77	15	Q90Q16	Q90Q16 human immun
126	35	79.5	215	5	Q26880	Q26880 trypanosoma	199	34	77.3	77	15	Q38442	Q38442 human immun
127	35	79.5	216	5	Q45932	Q45932 caenorhabdi	200	34	77.3	77	15	Q38436	Q38436 human immun
128	35	79.5	219	12	Q8QHR9	Q8QHR9 influenza a	201	34	77.3	77	15	Q90QG8	Q90QG8 human immun
129	35	79.5	249	12	Q92435	Q92435 bombix mori	202	34	77.3	78	15	Q91F15	Q91F15 human immun
130	35	79.5	267	11	Q9D255	Q9D255 mus musculu	203	34	77.3	78	15	Q38450	Q38450 human immun
131	35	79.5	320	6	Q95KV4	Q95KV4 bos taurus	204	34	77.3	78	15	Q38433	Q38433 human immun
132	35	79.5	347	6	Q95KV5	Q95KV5 bos taurus	205	34	77.3	78	15	Q38440	Q38440 human immun
133	35	79.5	354	4	Q9UQ56	Q9UQ56 homo sapien	206	34	77.3	78	15	Q38468	Q38468 human immun
134	35	79.5	374	10	Q39307	Q39307 brassica na	207	34	77.3	78	15	Q38449	Q38449 human immun
135	35	79.5	376	10	Q39306	Q39306 brassica na	208	34	77.3	78	15	Q90QC7	Q90QC7 human immun
136	35	79.5	378	11	Q99MW7	Q99MW7 mus musculu	209	34	77.3	78	15	Q38469	Q38469 human immun
137	35	79.5	379	4	Q95617	Q95617 homo sapien	210	34	77.3	78	15	Q38434	Q38434 human immun
138	35	79.5	382	13	Q7ZVT0	Q7ZVT0 brachydantio	211	34	77.3	78	15	Q38124	Q38124 human immun
139	35	79.5	407	17	Q29718	Q29718 archaeoglob	212	34	77.3	78	15	Q90QG4	Q90QG4 human immun
140	35	79.5	440	13	Q90816	Q90816 gallus gall	213	34	77.3	78	15	Q90QH6	Q90QH6 human immun
141	35	79.5	446	4	Q8N3Y0	Q8N3Y0 homo sapien	214	34	77.3	78	15	Q38437	Q38437 human immun
142	35	79.5	471	11	Q9D3R9	Q9D3R9 mus musculu	215	34	77.3	78	15	Q90QC9	Q90QC9 human immun
143	35	79.5	486	16	Q8DJN2	Q8DJN2 synechococc	216	34	77.3	79	15	Q38467	Q38467 human immun
144	35	79.5	504	10	Q9LH00	Q9LH00 arabidopsis	217	34	77.3	79	15	Q38441	Q38441 human immun
145	35	79.5	507	5	Q968L9	Q968L9 trypanosoma	218	34	77.3	79	15	Q38435	Q38435 human immun
146	35	79.5	516	10	Q9SLI0	Q9SLI0 arabidopsis	219	34	77.3	79	15	Q90QC5	Q90QC5 human immun
147	35	79.5	539	10	Q941P5	Q941P5 arabidopsis	220	34	77.3	79	15	Q90QES	Q90QES human immun
148	35	79.5	539	10	Q8L787	Q8L787 arabidopsis	221	34	77.3	79	15	Q38448	Q38448 human immun
149	35	79.5	558	3	Q13330	Q13330 emericella	222	34	77.3	79	15	Q90QD8	Q90QD8 human immun
150	35	79.5	609	17	Q8TXA4	Q8TXA4 methanopyru	223	34	77.3	79	15	Q38471	Q38471 human immun
151	35	79.5	776	11	Q8K1N1	Q8K1N1 mus musculu	224	34	77.3	80	15	Q38439	Q38439 human immun
152	35	79.5	784	11	Q8C605	Q8C605 mus musculu	225	34	77.3	80	15	Q90QH4	Q90QH4 human immun
153	35	79.5	784	11	Q8C5I6	Q8C5I6 mus musculu	226	34	77.3	80	15	Q90QD5	Q90QD5 human immun
154	35	79.5	803	11	Q9DC20	Q9DC20 mus musculu	227	34	77.3	81	15	Q90QI0	Q90QI0 human immun
155	35	79.5	810	11	Q8R3F3	Q8R3F3 mus musculu	228	34	77.3	81	15	Q38457	Q38457 human immun
156	35	79.5	922	11	Q8BZX2	Q8BZX2 mus musculu	229	34	77.3	82	15	Q90QJ2	Q90QJ2 human immun
157	35	79.5	1157	1	Q9HH42	Q9HH42 methanobact	230	34	77.3	82	15	Q38456	Q38456 human immun
158	35	79.5	1179	16	Q9RVV9	Q9RVV9 deinococcus	231	34	77.3	84	15	Q38477	Q38477 human immun
159	35	79.5	1186	9	Q80211	Q80211 methanobact	232	34	77.3	86	15	Q38289	Q38289 human immun
160	35	79.5	1399	4	Q75870	Q75870 homo sapien	233	34	77.3	86	15	Q38459	Q38459 human immun
161	35	79.5	1863	11	Q64605	Q64605 rattus norv	234	34	77.3	86	15	Q38517	Q38517 human immun
162	35	79.5	1894	11	Q64487	Q64487 mus musculu	235	34	77.3	86	15	Q38037	Q38037 human immun



236	34	77.3	86	15	038306	038306	human	immun	309	34	77.3	86	15	038458	038458	human	immun
237	34	77.3	86	15	038452	038452	human	immun	310	34	77.3	86	15	038278	038278	human	immun
238	34	77.3	86	15	038460	038460	human	immun	311	34	77.3	86	15	038325	038325	human	immun
239	34	77.3	86	15	038331	038331	human	immun	312	34	77.3	86	15	038248	038248	human	immun
240	34	77.3	86	15	038297	038297	human	immun	313	34	77.3	86	15	038308	038308	human	immun
241	34	77.3	86	15	038317	038317	human	immun	314	34	77.3	86	15	038290	038290	human	immun
242	34	77.3	86	15	038038	038038	human	immun	315	34	77.3	86	15	038536	038536	human	immun
243	34	77.3	86	15	038446	038446	human	immun	316	34	77.3	86	15	038304	038304	human	immun
244	34	77.3	86	15	038489	038489	human	immun	317	34	77.3	86	15	038268	038268	human	immun
245	34	77.3	86	15	038263	038263	human	immun	318	34	77.3	86	15	038301	038301	human	immun
246	34	77.3	86	15	038272	038272	human	immun	319	34	77.3	86	15	038275	038275	human	immun
247	34	77.3	86	15	038535	038535	human	immun	320	34	77.3	86	15	038462	038462	human	immun
248	34	77.3	86	15	038246	038246	human	immun	321	34	77.3	86	15	038497	038497	human	immun
249	34	77.3	86	15	038039	038039	human	immun	322	34	77.3	86	15	038334	038334	human	immun
250	34	77.3	86	15	038293	038293	human	immun	323	34	77.3	86	15	038287	038287	human	immun
251	34	77.3	86	15	038280	038280	human	immun	324	34	77.3	86	15	038264	038264	human	immun
252	34	77.3	86	15	038318	038318	human	immun	325	34	77.3	86	15	038245	038245	human	immun
253	34	77.3	86	15	038455	038455	human	immun	326	34	77.3	86	15	038283	038283	human	immun
254	34	77.3	86	15	038262	038262	human	immun	327	34	77.3	86	15	038533	038533	human	immun
255	34	77.3	86	15	038256	038256	human	immun	328	34	77.3	86	15	038252	038252	human	immun
256	34	77.3	86	15	038273	038273	human	immun	329	34	77.3	86	15	038323	038323	human	immun
257	34	77.3	86	15	038322	038322	human	immun	330	34	77.3	86	15	038277	038277	human	immun
258	34	77.3	86	15	038253	038253	human	immun	331	34	77.3	86	15	038324	038324	human	immun
259	34	77.3	86	15	038288	038288	human	immun	332	34	77.3	86	15	038305	038305	human	immun
260	34	77.3	86	15	038531	038531	human	immun	333	34	77.3	86	15	038254	038254	human	immun
261	34	77.3	86	15	038036	038036	human	immun	334	34	77.3	86	15	038247	038247	human	immun
262	34	77.3	86	15	038332	038332	human	immun	335	34	77.3	86	15	038309	038309	human	immun
263	34	77.3	86	15	038270	038270	human	immun	336	34	77.3	86	15	038282	038282	human	immun
264	34	77.3	86	15	038035	038035	human	immun	337	34	77.3	86	15	038515	038515	human	immun
265	34	77.3	86	15	038281	038281	human	immun	338	34	77.3	86	15	038444	038444	human	immun
266	34	77.3	86	15	038330	038330	human	immun	339	34	77.3	86	15	038426	038426	human	immun
267	34	77.3	86	15	038327	038327	human	immun	340	34	77.3	86	15	038481	038481	human	immun
268	34	77.3	86	15	038250	038250	human	immun	341	34	77.3	86	15	038500	038500	human	immun
269	34	77.3	86	15	038328	038328	human	immun	342	34	77.3	86	15	038496	038496	human	immun
270	34	77.3	86	15	038326	038326	human	immun	343	34	77.3	86	15	038427	038427	human	immun
271	34	77.3	86	15	038298	038298	human	immun	344	34	77.3	86	15	038461	038461	human	immun
272	34	77.3	86	15	038451	038451	human	immun	345	34	77.3	86	15	038425	038425	human	immun
273	34	77.3	86	15	038529	038529	human	immun	346	34	77.3	86	15	038465	038465	human	immun
274	34	77.3	86	15	038315	038315	human	immun	347	34	77.3	86	15	038513	038513	human	immun
275	34	77.3	86	15	038463	038463	human	immun	348	34	77.3	86	15	038484	038484	human	immun
276	34	77.3	86	15	038251	038251	human	immun	349	34	77.3	86	15	038512	038512	human	immun
277	34	77.3	86	15	038320	038320	human	immun	350	34	77.3	86	15	038447	038447	human	immun
278	34	77.3	86	15	038279	038279	human	immun	351	34	77.3	86	15	038502	038502	human	immun
279	34	77.3	86	15	038302	038302	human	immun	352	34	77.3	86	15	038464	038464	human	immun
280	34	77.3	86	15	038333	038333	human	immun	353	34	77.3	86	15	038528	038528	human	immun
281	34	77.3	86	15	038259	038259	human	immun	354	34	77.3	86	15	038474	038474	human	immun
282	34	77.3	86	15	038313	038313	human	immun	355	34	77.3	86	15	038486	038486	human	immun
283	34	77.3	86	15	038490	038490	human	immun	356	34	77.3	86	15	038501	038501	human	immun
284	34	77.3	86	15	038291	038291	human	immun	357	34	77.3	86	15	038480	038480	human	immun
285	34	77.3	86	15	038261	038261	human	immun	358	34	77.3	86	15	038472	038472	human	immun
286	34	77.3	86	15	038307	038307	human	immun	359	34	77.3	86	15	038479	038479	human	immun
287	34	77.3	86	15	038255	038255	human	immun	360	34	77.3	86	15	038510	038510	human	immun
288	34	77.3	86	15	038303	038303	human	immun	361	34	77.3	86	15	038470	038470	human	immun
289	34	77.3	86	15	038319	038319	human	immun	362	34	77.3	86	15	038493	038493	human	immun
290	34	77.3	86	15	038260	038260	human	immun	363	34	77.3	86	15	038482	038482	human	immun
291	34	77.3	86	15	038505	038505	human	immun	364	34	77.3	86	15	038538	038538	human	immun
292	34	77.3	86	15	038454	038454	human	immun	365	34	77.3	86	15	038503	038503	human	immun
293	34	77.3	86	15	038271	038271	human	immun	366	34	77.3	86	15	038518	038518	human	immun
294	34	77.3	86	15	038296	038296	human	immun	367	34	77.3	86	15	038537	038537	human	immun
295	34	77.3	86	15	038249	038249	human	immun	368	34	77.3	86	15	038494	038494	human	immun
296	34	77.3	86	15	038258	038258	human	immun	369	34	77.3	86	15	038514	038514	human	immun
297	34	77.3	86	15	038265	038265	human	immun	370	34	77.3	86	15	038509	038509	human	immun
298	34	77.3	86	15	038519	038519	human	immun	371	34	77.3	86	15	038508	038508	human	immun
299	34	77.3	86	15	038314	038314	human	immun	372	34	77.3	86	15	038485	038485	human	immun
300	34	77.3	86	15	038266	038266	human	immun	373	34	77.3	86	15	038537	038537	human	immun
301	34	77.3	86	15	038312	038312	human	immun	374	34	77.3	86	15	038504	038504	human	immun
302	34	77.3	86	15	038453	038453	human	immun	375	34	77.3	86	15	038483	038483	human	immun
303	34	77.3	86	15	038311	038311	human	immun	376	34	77.3	86	15	038507	038507	human	immun
304	34	77.3	86	15	038532	038532	human	immun	377	34	77.3	86	15	038466	038466	human	immun
305	34	77.3	86	15	038321	038321	human	immun	378	34	77.3	86	15	038475	038475	human	immun
306	34	77.3	86	15	038276	038276	human	immun	379	34	77.3	86	15	038473	038473	human	immun
307	34	77.3	86	15	038274	038274	human	immun	380	34	77.3	86	15	038445	038445	human	immun
308	34	77.3	86	15	038267	038267	human	immun	381	34	77.3	86	15	038511	038511	human	immun

382	34	77.3	88	15	038476	038476	human	immun
383	34	77.3	88	15	038495	038495	human	immun
384	34	77.3	88	15	038498	038498	human	immun
385	34	77.3	88	15	038499	038499	human	immun
386	34	77.3	88	15	038506	038506	human	immun
387	34	77.3	88	15	038506	038506	human	immun
388	34	77.3	88	15	038506	038506	human	immun
389	34	77.3	88	15	038506	038506	human	immun
390	34	77.3	88	15	038506	038506	human	immun
391	34	77.3	88	15	038506	038506	human	immun
392	34	77.3	88	15	038506	038506	human	immun
393	34	77.3	88	15	038506	038506	human	immun
394	34	77.3	88	15	038506	038506	human	immun
395	34	77.3	88	15	038506	038506	human	immun
396	34	77.3	88	15	038506	038506	human	immun
397	34	77.3	88	15	038506	038506	human	immun
398	34	77.3	88	15	038506	038506	human	immun
399	34	77.3	88	15	038506	038506	human	immun
400	34	77.3	88	15	038506	038506	human	immun
401	34	77.3	88	15	038506	038506	human	immun
402	34	77.3	88	15	038506	038506	human	immun
403	34	77.3	88	15	038506	038506	human	immun
404	34	77.3	88	15	038506	038506	human	immun
405	34	77.3	88	15	038506	038506	human	immun
406	34	77.3	88	15	038506	038506	human	immun
407	34	77.3	88	15	038506	038506	human	immun
408	34	77.3	88	15	038506	038506	human	immun
409	34	77.3	88	15	038506	038506	human	immun
410	34	77.3	88	15	038506	038506	human	immun
411	34	77.3	88	15	038506	038506	human	immun
412	34	77.3	88	15	038506	038506	human	immun
413	34	77.3	88	15	038506	038506	human	immun
414	34	77.3	88	15	038506	038506	human	immun
415	34	77.3	88	15	038506	038506	human	immun
416	34	77.3	88	15	038506	038506	human	immun
417	34	77.3	88	15	038506	038506	human	immun
418	34	77.3	88	15	038506	038506	human	immun
419	34	77.3	88	15	038506	038506	human	immun
420	34	77.3	88	15	038506	038506	human	immun
421	34	77.3	88	15	038506	038506	human	immun
422	34	77.3	88	15	038506	038506	human	immun
423	34	77.3	88	15	038506	038506	human	immun
424	34	77.3	88	15	038506	038506	human	immun
425	34	77.3	88	15	038506	038506	human	immun
426	34	77.3	88	15	038506	038506	human	immun
427	34	77.3	88	15	038506	038506	human	immun
428	34	77.3	88	15	038506	038506	human	immun
429	34	77.3	88	15	038506	038506	human	immun
430	34	77.3	88	15	038506	038506	human	immun
431	34	77.3	88	15	038506	038506	human	immun
432	34	77.3	88	15	038506	038506	human	immun
433	34	77.3	88	15	038506	038506	human	immun
434	34	77.3	88	15	038506	038506	human	immun
435	34	77.3	88	15	038506	038506	human	immun
436	34	77.3	88	15	038506	038506	human	immun
437	34	77.3	88	15	038506	038506	human	immun
438	34	77.3	88	15	038506	038506	human	immun
439	34	77.3	88	15	038506	038506	human	immun
440	34	77.3	88	15	038506	038506	human	immun
441	34	77.3	88	15	038506	038506	human	immun
442	34	77.3	88	15	038506	038506	human	immun
443	34	77.3	88	15	038506	038506	human	immun
444	34	77.3	88	15	038506	038506	human	immun
445	34	77.3	88	15	038506	038506	human	immun
446	34	77.3	88	15	038506	038506	human	immun
447	34	77.3	88	15	038506	038506	human	immun
448	34	77.3	88	15	038506	038506	human	immun
449	34	77.3	88	15	038506	038506	human	immun
450	34	77.3	88	15	038506	038506	human	immun
451	34	77.3	88	15	038506	038506	human	immun
452	34	77.3	88	15	038506	038506	human	immun
453	34	77.3	88	15	038506	038506	human	immun
454	34	77.3	88	15	038506	038506	human	immun

455	34	77.3	94	15	Q9DYS0	Q9DYS0	human	immun
456	34	77.3	94	15	Q9DZ07	Q9DZ07	human	immun
457	34	77.3	94	15	Q9DZ01	Q9DZ01	human	immun
458	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
459	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
460	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
461	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
462	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
463	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
464	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
465	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
466	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
467	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
468	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
469	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
470	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
471	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
472	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
473	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
474	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
475	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
476	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
477	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
478	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
479	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
480	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
481	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
482	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
483	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
484	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
485	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
486	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
487	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
488	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
489	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
490	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
491	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
492	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
493	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
494	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
495	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
496	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
497	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
498	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
499	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun
500	34	77.3	94	15	Q9DZ75	Q9DZ75	human	immun

## ALIGNMENTS

## RESULT 1

ID	Q8H611	PRELIMINARY;	PRT;	566 AA.
AC	Q8H611;			
DT	01-MAR-2003 (TREMBLrel. 23, Created)			
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)			
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)			
DE	Puative MURAZC.			
GN	ZMR5072.18.			
OS	Zea mays (Maize).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;			
OC	PACCAD clade; Panicoideae; Andropogoneae; Zea.			
OX	NCBI_TaxID=4577;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. BSSS53;			
RX	MEDLINE=2253764; PubMed=12368247;			
RA	Song R., Laca V., Messing J.;			
RT	"Mosaic Organization of Orthologous Sequences in Grass Genomes.";			
RL	Genome Res. 12:1549-1555(2002).			
DR	EMBL; AF528565; AAN40034.1; -.			

DR InterPro: IPR001207; Trnspase\_mutatr.  
 DR InterPro: IPR006564; Znf\_PWZ.  
 DR InterPro: IPR007527; Znf\_SWIM.  
 DR Pfam: PF04434; SWIM; 1.  
 DR SMART: SM00575; Znf\_PWZ; 1.  
 DR PROSITE: PS01007; TRANSPOSASE\_MUTATOR; 1.  
 SQ SEQUENCE 566 AA; 65236 MW; AF06A330CA3E8C36 CRC64;

Query Match 93.2%; Score 41; DB 10; Length 566;  
 Best Local Similarity 87.5%; Pred. No. 38;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KPSSPPE 8  
 Db 542 RPSSPPE 549

## RESULT 2

Q8R289 PRELIMINARY; PRT; 390 AA.  
 AC Q8R289;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DE OSUNB0063G05.9 protein.  
 GN OSUNB0063G05.9.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC  
 RT clone:OSUNB0063G05.9";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL:AF003760; BAB90500.1; -.  
 DR Gramene; Q8R289; -.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0016462; F:pyrophosphatase activity; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR InterPro: IPR008162; Pyrophosphatase.  
 DR PROSITE: PS00387; PASE; 1.  
 SQ SEQUENCE 390 AA; 43198 MW; 23CE65ED8A3745A0 CRC64;

Query Match 88.6%; Score 39; DB 10; Length 390;  
 Best Local Similarity 100.0%; Pred. No. 58;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 PSSPPE 8  
 Db 231 PSSPPE 237

## RESULT 3

Q8L775 PRELIMINARY; PRT; 802 AA.  
 AC Q8L775;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE At1g07990/T6D22.5.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC euroids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Shim P., Chen H., Cheuk R., Kim C.J., Banh J., Bowser L.,  
 RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,

RA Iehida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,  
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,  
 RA Palm C.C., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
 RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,  
 RA Shinozaki K., Davis R.W., Theologis A., Becker J.R.;  
 RT "Arabidopsis cDNA clones."  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.  
 RA Cheuk R., Chen H., Kim C.J., Shim P., Bowser L., Carninci P.,  
 RA Chan M.M., Chang C.H., Dale J.M., Hayashizaki Y., Heuan V.W.,  
 RA Iehida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,  
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,  
 RA Palm C.C., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
 RA Tang C.C., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G., Yuan S.,  
 RA Shinozaki K., Davis R.W., Theologis A., Becker J.R.;  
 RT "Arabidopsis ORF clones."  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL:AY127028; AAM83251.1; -.  
 DR EMBL:BT004535; AAO42781.1; -.  
 DR InterPro: IPR007587; SAPS.  
 DR Pfam: PF04499; SAPS; 1.  
 SQ SEQUENCE 802 AA; 89588 MW; 705F6276049B4F32 CRC64;

Query Match 88.6%; Score 39; DB 10; Length 802;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KPSSPPE 7  
 Db 202 KPSSPPE 208

## RESULT 4

Q9JN08 PRELIMINARY; PRT; 990 AA.  
 AC Q9JN08;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE T6D22.8.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC euroids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Shim P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,  
 RA Kim C., Altafi H., Bei Q., Chin C., Chou J., Choi B., Conn L.,  
 RA Conway A., Gonzalez A., Hansen N., Howling B., Koo T., Lam B., Lee J.,  
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharly N., Nguyen M.,  
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thavert A.,  
 RA Toriumi M., Vayberg M., Yu G., Federpiel N.A., Theologis A.,  
 RA Becker J.R.;  
 RT "Genomic sequence for Arabidopsis thaliana BAC T6D22 from chromosome  
 RT 1."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Becker J.R.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Becker J.R.;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Shim P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
 RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chou J., Choi E.,  
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howling B., Koo T., Lam B.,  
 RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharly N.,  
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,

RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,  
 RA Theologis A., Ecker J.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC026875; AAF79824.1; -.  
 DR PIR: A86215; A86215.  
 DR InterPro: IPR007587; SAPS.  
 DR Pfam: PF04499; SAPS; 1.  
 SQ SEQUENCE 990 AA; 110996 MW; E8687D5C04D298E CRC64;

Query Match 86.4%; Score 39; DB 10; Length 990;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 7  
 Db 381 KPSSPPE 387

## RESULT 5

ID 038534 PRELIMINARY; PRT; 87 AA.  
 AC 038534;  
 DT 01-JUN-1998 (TEMBLrel. 05, Created)  
 DT 01-JUN-1998 (TEMBLrel. 05, Last sequence update)  
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
 DE Gag protein (Fragment).  
 GN GAG.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 NC NCB1\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97404676; Pubmed=9261388;  
 RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,  
 RA Vasudevachari M.B., Salzman N.P.;  
 RT "Drug resistance during indinavir therapy is caused by mutations in  
 the protease gene and in its Gag substrate cleavage sites.";  
 RL J. Virol. 71:6662-6670(1997).  
 DR EMBL: AF024433; AAB8307.1; -.  
 DR HSP: P05888; IAAF.  
 DR GO: GO:0019012; Cytolysis; IEA.  
 DR GO: GO:0003676; F-nucleic acid binding; IEA.  
 DR InterPro: IPR001878; ZnF\_CCHC.  
 DR Pfam: PF00098; Zf-CCHC; 1.  
 DR PRINTS: PR00939; C2HCZNFINGER.  
 DR SMART: SM00343; ZnF\_CCHC; 1.  
 DR PROSITE: PS50158; ZF\_CCHC; 1.  
 KM Core protein; Polyprotein.  
 FT NON\_TER 1  
 FT NON\_TER 87  
 SQ SEQUENCE 87 AA; 10037 MW; C0525F43D8210973 CRC64;

Query Match 86.4%; Score 38; DB 15; Length 87;  
 Best Local Similarity 75.0%; Pred. No. 19;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
 Db 53 KPAPPE 60

## RESULT 6

ID 099J03 PRELIMINARY; PRT; 215 AA.  
 AC 099J03;  
 DT 01-JUN-2001 (TEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
 DE Similar to hypothetical protein FLJ20837.  
 GN SLC30A6.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

QY 1 KPSSPPE 8  
 Db 53 KPAPPE 60

OX NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC005753; AAH05753.1; -.  
 DR MGD: MGI:2386741; SLC30A6.  
 DR GO: GO:0005385; F-ionic ion transporter activity; IDA.  
 DR GO: GO:0006829; P-golgi to endosome transport; IDA.  
 DR GO: GO:0006829; P-zinc ion transport; IDA.  
 DR InterPro: IPR002524; Cation\_efflux.  
 DR Pfam: PF01545; Cation\_efflux; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 215 AA; 23529 MW; 9B58AF6C62B3872 CRC64;

Query Match 86.4%; Score 38; DB 11; Length 215;  
 Best Local Similarity 87.5%; Pred. No. 48;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
 Db 134 KPSSPPE 141

## RESULT 7

ID 08N5C9 PRELIMINARY; PRT; 216 AA.  
 AC 08N5C9;  
 DT 01-OCT-2002 (TEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)  
 DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)  
 DE Similar to hypothetical protein MGC11963.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NC NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Eye;  
 RA Strausberg R.;  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC032525; AAH32525.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 216 AA; 23562 MW; DD7E7F862B78F8FB3 CRC64;

Query Match 86.4%; Score 38; DB 4; Length 216;  
 Best Local Similarity 87.5%; Pred. No. 48;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
 Db 134 KPSSPPE 141

## RESULT 8

ID 08R422 PRELIMINARY; PRT; 460 AA.  
 AC 08R422;  
 DT 01-JUN-2002 (TEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)  
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
 DE Zinc transporter-like 3 protein.  
 GN SLC30A6.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RA Zhu W., Mager S.;  
 RL "Cloning of new mammalian zinc transporter-like genes.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

QY 1 KPSSPPE 8  
 Db 134 KPSSPPE 141

DR EMBL: AF233346; AAL8317.1; -.  
 DR MGD; MGI:2386741; SLC30A6.  
 DR GO; GO:0005385; P:zinc ion transporter activity; IDA.  
 DR GO; GO:0006895; P:golgi to endosome transport; IDA.  
 DR GO; GO:0006829; P:zinc ion transport; IDA.  
 DR InterPro; IPR002524; Cation\_efflux.  
 DR Pfam; PF01545; Cation\_efflux; 1.  
 DR TIGRPFAM; TIGR01297; CDF; 1.  
 SQ SEQUENCE 460 AA; 51015 MW; A05DD3116B0BC051 CRC64;

Query Match 86.4%; Score 38; DB 11; Length 460;  
 Best Local Similarity 87.5%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 KPSSPPEE 8  
 DB 379 KPSSPPEE 386

RESULT 9  
 ID Q8K4H6 PRELIMINARY; PRT; 460 AA.

AC Q8K4H6; 01-OCT-2002 (T-EMBLrel. 22, Created)  
 DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
 DE Zinc transporter 6.  
 GN SLC30A6 OR ZNT6.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CS7BL/6J;  
 RX Huang L., Kirschke C.P., Gilechier J.;  
 RT "Functional Characterization of a Novel Mammalian Zinc Transporter,  
 Znt6.";  
 RL J. Biol. Chem. 277:26389-26395(2002).  
 DR EMBL; AF395840; AAM27917.1; -.  
 DR MGD; MGI:2386741; SLC30A6.  
 DR GO; GO:0005385; P:zinc ion transporter activity; IDA.  
 DR GO; GO:0006895; P:golgi to endosome transport; IDA.  
 DR GO; GO:0006829; P:zinc ion transport; IDA.  
 DR InterPro; IPR002524; Cation\_efflux.  
 DR Pfam; PF01545; Cation\_efflux; 1.  
 DR TIGRPFAM; TIGR01297; CDF; 1.  
 SQ SEQUENCE 460 AA; 50894 MW; B61A2F0E0AD6921B CRC64;

Query Match 86.4%; Score 38; DB 11; Length 460;  
 Best Local Similarity 87.5%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 KPSSPPEE 8  
 DB 379 KPSSPPEE 386

RESULT 10  
 ID Q8BJM5 PRELIMINARY; PRT; 460 AA.

AC Q8BJM5; 01-MAR-2003 (T-EMBLrel. 23, Created)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
 DE CDNA FLJ31101 FIS.  
 GN SLC30A6.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCB1\_TaxID=10090;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=CS7BL/6J; TISSUE=Body;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium.  
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK082807; BAC38680.1; -.  
 DR MGD; MGI:2386741; SLC30A6.  
 DR GO; GO:0005385; P:zinc ion transporter activity; IDA.  
 DR GO; GO:0006895; P:golgi to endosome transport; IDA.  
 DR GO; GO:0006829; P:zinc ion transport; IDA.  
 DR InterPro; IPR002524; Cation\_efflux.  
 DR Pfam; PF01545; Cation\_efflux; 1.  
 DR TIGRPFAM; TIGR01297; CDF; 1.  
 SQ SEQUENCE 460 AA; 51026 MW; A3B910CB96D16015 CRC64;

Query Match 86.4%; Score 38; DB 11; Length 460;  
 Best Local Similarity 87.5%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 KPSSPPEE 8  
 DB 379 KPSSPPEE 386

RESULT 11  
 ID Q96NC3 PRELIMINARY; PRT; 461 AA.

AC Q96NC3; 01-DEC-2001 (T-EMBLrel. 19, Created)  
 DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)  
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)  
 DE Hypothetical protein FLJ31101.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,  
 RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,  
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
 RA Kawai-Hiro Y., Saigo K., Nishikawa T., Kimura K., Yamashita H.,  
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagaesuma M.,  
 RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,  
 RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,  
 RA Isegaki T.;  
 RT "NEDO human cDNA sequencing project.";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK055663; BAB70980.1; -.  
 DR Gene; HGNC:19305; SLC30A6.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0008324; P:cation transporter activity; IEA.  
 DR GO; GO:0006812; P:cation transport; IEA.  
 DR InterPro; IPR002524; Cation\_efflux.  
 DR Pfam; PF01545; Cation\_efflux; 1.  
 DR TIGRPFAM; TIGR01297; CDF; 1.  
 KM Hypothetical protein.  
 SQ SEQUENCE 461 AA; 51115 MW; 7926F6888EF1F1F9 CRC64;

Query Match 86.4%; Score 38; DB 4; Length 461;  
 Best Local Similarity 87.5%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 KPSSPPEE 8  
 DB 379 KPSSPPEE 386

RESULT 12  
 Q9LUT9

ID Q9L079 PRELIMINARY; PRT; 607 AA.  
AC Q9L079;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)  
DE Db|AA|21150.1.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Columbia;  
RX MEDLINE=20181125; PubMed=10718197;  
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,  
RA Tabata S.;  
RT "structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence  
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC  
RT clones."  
RL DNA Res. 7:31-63(2000).  
DR EMBL; AB023042; BAA97357.1; -.  
SQ SEQUENCE 607 AA; 67287 MW; C67FE8C2F21D9317 CRC64;

Query Match 86.4%; Score 38; DB 10; Length 607;  
Best Local Similarity 75.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8  
Db 240 KPSSPPEQ 247

RESULT 13  
Q872R5 PRELIMINARY; PRT; 1351 AA.  
AC Q872R5;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)  
DE Related to RNA-directed RNA polymerase.  
GN B13B3.100.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,  
RA Nyakatura G., Mewes H.W., Mannhaupt G.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA German Neurospora genome project.  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BX284762; CAD70515.1; -.  
DR GO; GO:0016329; F:apoptosis regulator activity; IEA.  
DR GO; GO:000368; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0006915; P:apoptosis; IEA.  
DR InterPro; IPR000712; Bcl2\_BH.  
DR InterPro; IPR007855; RDRP.  
DR InterPro; IPR003124; WH2.  
DR Pfam; PF05183; RDRP.1.  
DR PROSITE; PS01259; BH3.1.  
KM RNA-directed RNA polymerase.  
SQ SEQUENCE 1351 AA; 151975 MW; 0C784B99CDC66AB6 CRC64;

Query Match 86.4%; Score 38; DB 3; Length 1351;  
Best Local Similarity 75.0%; Pred. No. 3.1e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8

Db 1132 RPSPPPEE 1139

RESULT 14  
Q99308 PRELIMINARY; PRT; 52 AA.  
AC Q99308;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, last annotation update)  
DE Gag protein (Fragment).  
GN GAG.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B497;  
RX MEDLINE=21136890; PubMed=11238855;  
RA Bleiber G., Munoz M., Cluffi A., Meylan P., Telenti A.;  
RT "Individual Contributions of Mutant Protease and Reverse Transcriptase  
RT to Viral Infectivity, Replication, and Protein Maturation of  
RT Antiretroviral Drug-Resistant Human Immunodeficiency Virus Type 1."  
RL J. Virol. 75:3291-3300(2001).  
DR EMBL; AF316838; AAK21082.1; -.  
FT NON\_TER  
SQ SEQUENCE 52 AA; 5704 MW; 1312CE4A8CF0D75 CRC64;

Query Match 84.1%; Score 37; DB 15; Length 52;  
Best Local Similarity 75.0%; Pred. No. 17;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8  
Db 6 EPSAPPEE 13

RESULT 15  
Q99350 PRELIMINARY; PRT; 52 AA.  
AC Q99350;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, last annotation update)  
DE Gag protein (Fragment).  
GN GAG.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B495;  
RX MEDLINE=21136890; PubMed=11238855;  
RA Bleiber G., Munoz M., Cluffi A., Meylan P., Telenti A.;  
RT "Individual Contributions of Mutant Protease and Reverse Transcriptase  
RT to Viral Infectivity, Replication, and Protein Maturation of  
RT Antiretroviral Drug-Resistant Human Immunodeficiency Virus Type 1."  
RL J. Virol. 75:3291-3300(2001).  
DR EMBL; AF316831; AAK21070.1; -.  
FT NON\_TER  
SQ SEQUENCE 52 AA; 5719 MW; 47C0789171E09693 CRC64;

Query Match 84.1%; Score 37; DB 15; Length 52;  
Best Local Similarity 75.0%; Pred. No. 17;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8  
Db 6 EPSAPPEE 13

RESULT 16

Q91F19  
ID Q91F19 PRELIMINARY; PRT; 67 AA.  
AC Q91F19;  
DT 01-OCT-2000 (TEMBLrel. 15, Created)  
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)  
DE 01-OCT-2002 (TEMBLrel. 22, Last annotation update)  
GN Gag protein (Fragment).  
OS Human immunodeficiency virus 1.  
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1145;  
RX MEDLINE=21443958; PubMed=11559796;  
RA Peters S., Munoz M., Yerly S., Sanchez-Merino V., Lopez-Galindez C.,  
RA Perrin L., Larder B., Cmaro D., Fakan S., Maylan P., Telenti A.;  
RT "Resistance to Nucleoside Analog Reverse Transcriptase Inhibitors  
Mediated by Human Immunodeficiency Virus Type 1 p6 Protein.";  
RL J. Virol. 75:9644-9653(2001).  
DR EMBL; AF282952; AAF87816.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 67 AA; 7470 MW; 810C3A06512B9A65 CRC64;

Query Match 84.1%; Score 37; DB 15; Length 67;  
Best Local Similarity 75.0%; Pred. No. 22;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
Db 22 EPSAPPE 29

RESULT 17  
Q9DYK6  
ID Q9DYK6 PRELIMINARY; PRT; 68 AA.  
AC Q9DYK6;  
DT 01-MAR-2001 (TEMBLrel. 16, Created)  
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)  
DT 01-OCT-2002 (TEMBLrel. 22, Last annotation update)  
DE Gag protein (Fragment).  
GN GAG.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=2016;  
RA Telenti A., Martinez R., Peters S.;  
RT "Natural mutants of HIV-1 subtype B frameshift gag p1 stemloop.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF293421; AAG15244.1; -.  
FT NON\_TER 1  
FT VARIANT 9  
SQ SEQUENCE 68 AA; 7546 MW; 03E7B81A6F742393 CRC64;

Query Match 84.1%; Score 37; DB 15; Length 68;  
Best Local Similarity 75.0%; Pred. No. 22;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
Db 22 EPSAPPE 29

RESULT 18  
Q91F25  
ID Q91F25 PRELIMINARY; PRT; 68 AA.  
AC Q91F25;  
DT 01-OCT-2000 (TEMBLrel. 15, Created)  
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)  
DT 01-OCT-2002 (TEMBLrel. 22, Last annotation update)  
DE Gag protein (Fragment).

GN POL.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1186;  
RX MEDLINE=21443958; PubMed=11559796;  
RA Peters S., Munoz M., Yerly S., Sanchez-Merino V., Lopez-Galindez C.,  
RA Perrin L., Larder B., Cmaro D., Fakan S., Maylan P., Telenti A.;  
RT "Resistance to Nucleoside Analog Reverse Transcriptase Inhibitors  
Mediated by Human Immunodeficiency Virus Type 1 p6 Protein.";  
RL J. Virol. 75:9644-9653(2001).  
DR EMBL; AF282959; AAF87810.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 68 AA; 7585 MW; 8F1F542EB31A3D01 CRC64;

Query Match 84.1%; Score 37; DB 15; Length 68;  
Best Local Similarity 75.0%; Pred. No. 22;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
Db 22 EPSAPPE 29

RESULT 19  
Q38224  
ID Q38224 PRELIMINARY; PRT; 73 AA.  
AC Q38224;  
DT 01-JAN-1998 (TEMBLrel. 05, Created)  
DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)  
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
DE Gag protein (Fragment).  
GN GAG.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97404676; PubMed=9261388;  
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,  
RA Vasudevachari M.B., Salzman N.P.;  
RT "Drug resistance during indinavir therapy is caused by mutations in  
the protease gene and in its Gag substrate cleavage sites.";  
RL J. Virol. 71:6662-6670(1997).  
DR EMBL; AF024121; AAB82997.1; -.  
DR GO; GO:0019012; C:violation; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR SMART; SM00343; Znf\_C2HC; 1.  
KW Core protein; Polyprotein.  
FT NON\_TER 1  
FT NON\_TER 73  
SQ SEQUENCE 73 AA; 8224 MW; 2EA56BD672D4153 CRC64;

Query Match 84.1%; Score 37; DB 15; Length 73;  
Best Local Similarity 75.0%; Pred. No. 24;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
Db 39 EPSAPPE 46

RESULT 20  
Q38232  
ID Q38232 PRELIMINARY; PRT; 75 AA.  
AC Q38232;  
DT 01-JAN-1998 (TEMBLrel. 05, Created)  
DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)  
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
DE Gag protein (Fragment).

GN GAG.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97404676; PubMed=9261388;  
 RA Zhang Y.M., Imamiuchi H., Imamiuchi T., Lane H.C., Falloon J.,  
 RA Vasudevarachari M.B., Salzman N.P.;  
 RT "Drug resistance during indinavir therapy is caused by mutations in  
 the protease gene and in its Gag substrate cleavage sites.";  
 RL J. Virol. 71:6662-6670(1997).  
 DR EMBL: AF024129; AAB83005.1; -.  
 DR GO: GO:0019012; C:viral; IEA.  
 DR GO: GO:0003676; F:nucleic acid binding; IEA.  
 DR InterPro: IPR001878; Znf\_CCHC.  
 DR SMART: SM00343; Znf\_CCHC; 1.  
 KM Core protein; Polyprotein.  
 FT NON\_TER 1  
 FT SEQUENCE 75 AA; 8504 MW; 7F9E6A43622C92BF CRC64;  
 SQ

Query Match 84.1%; Score 37; DB 15; Length 75;  
 Best Local Similarity 75.0%; Pred. No. 25;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSPPEE 8  
 Db 39 EPSAPPEE 46

## RESULT 21

Q90QC3 PRELIMINARY; PRT; 76 AA.

ID Q90QC3  
 AC Q90QC3;  
 DT 01-DEC-2001 (TEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)  
 DE 01-OCT-2002 (TEMBLrel. 22, Last annotation update)  
 DR Gag protein (Fragment).  
 GN GAG.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=1248068; PubMed=11350662;  
 RA Kaufmann G.R., Suzuki K., Cunningham P., Mukai M., Kondo M.,  
 RA Imai M., Zaunders J., Cooper D.A.;  
 RT "Impact of HIV Type 1 Protease, Reverse Transcriptase, Cleavage Site,  
 and p6 Mutations on the Virological Response to Quadruple Therapy with  
 Zidovudine, Zalcitabine, Didanosine, and Zalcitabine.";  
 RT Sequinavir, Ritonavir, and Two Nucleoside Analogs.";  
 RL AIDS Res. Hum. Retroviruses 17:487-497(2001).  
 DR EMBL: AF323242; AK66729.1; -.  
 FT NON\_TER 1  
 FT SEQUENCE 76 AA; 8558 MW; 8BD09026451B1B90 CRC64;  
 SQ

Query Match 84.1%; Score 37; DB 15; Length 76;  
 Best Local Similarity 75.0%; Pred. No. 25;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSPPEE 8  
 Db 30 EPSAPPEE 37

## RESULT 22

Q90QH2 PRELIMINARY; PRT; 91 AA.

ID Q90QH2  
 AC Q90QH2;  
 DT 01-DEC-2001 (TEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
 DE Gag protein (Fragment).

GN GAG.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=1248068; PubMed=11350662;  
 RA Kaufmann G.R., Suzuki K., Cunningham P., Mukai M., Kondo M.,  
 RA Imai M., Zaunders J., Cooper D.A.;  
 RT "Impact of HIV Type 1 Protease, Reverse Transcriptase, Cleavage Site,  
 and p6 Mutations on the Virological Response to Quadruple Therapy with  
 Zidovudine, Zalcitabine, Didanosine, and Zalcitabine.";  
 RL AIDS Res. Hum. Retroviruses 17:487-497(2001).  
 DR EMBL: AF323242; AK66680.1; -.  
 DR GO: GO:0019012; C:viral; IEA.  
 DR GO: GO:0003676; F:nucleic acid binding; IEA.  
 DR InterPro: IPR001878; Znf\_CCHC.  
 DR Pfam: PF00098; Znf\_CCHC; 1.  
 DR PRINTS: PR00939; C2HCZNFINGER.  
 DR SMART: SM00343; Znf\_CCHC; 1.  
 DR PROSITE: PS50158; Znf\_CCHC; 1.  
 KM Core protein; Polyprotein.  
 FT NON\_TER 1  
 FT SEQUENCE 91 AA; 10233 MW; 30031F286DD006B1 CRC64;  
 SQ

Query Match 84.1%; Score 37; DB 15; Length 91;  
 Best Local Similarity 75.0%; Pred. No. 30;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSPPEE 8  
 Db 45 EPSAPPEE 52

## RESULT 23

Q9DYQ4 PRELIMINARY; PRT; 94 AA.

ID Q9DYQ4  
 AC Q9DYQ4;  
 DT 01-MAR-2001 (TEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)  
 DE 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
 DR Gag polyprotein (Fragment).  
 GN GAG.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20461476; PubMed=11005867;  
 RA Martinez-Picado J., DePaquale M.P., Karason N., Hanna G.J.,  
 RA Wong J., Finzi D., Rosenberg E., Gunthard H.F., Sutton L., Savara A.,  
 RA Petropoulos C.J., Hellmann N., Walker B.D., Richman D.D.,  
 RA Siliciano R., D'Aguiar R.T.;  
 RT "Antiretroviral resistance during successful therapy of HIV type 1  
 infection.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 97:10948-10953(2000).  
 DR EMBL: AF292888; AAG25536.1; -.  
 DR HSSP: P05888; IAAF.  
 DR GO: GO:0019012; C:viral; IEA.  
 DR GO: GO:0003676; F:nucleic acid binding; IEA.  
 DR InterPro: IPR001878; Znf\_CCHC.  
 DR Pfam: PF00098; Znf\_CCHC; 1.  
 DR PRINTS: PR00939; C2HCZNFINGER.  
 DR SMART: SM00343; Znf\_CCHC; 1.  
 DR PROSITE: PS50158; Znf\_CCHC; 1.  
 KM Core protein; Polyprotein.  
 FT NON\_TER 1  
 FT SEQUENCE 94 AA; 10642 MW; FDBF29CE4646B8E CRC64;  
 SQ

Query Match 84.1%; Score 37; DB 15; Length 94;  
 Best Local Similarity 75.0%; Pred. No. 31;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;



```

QY      1 KPSPPEE 8
      :||:||||
Db      48 EPSAPPEE 55

RESULT 24
09DY01      PRELIMINARY;      PRT;      94 AA.
AC      09DY01;
DT      01-MAR-2001 (TEMBLrel. 16, Created)
DT      01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DE      01-OCT-2003 (TEMBLrel. 25, Last annotation update)
GN      Gag polyprotein (Fragment).
OS      Human immunodeficiency virus 1.
OC      Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX      NCBI_TaxID=11676;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=20461476; PubMed=11005867;
RA      Martinez-Picado J., DePasquale M.P., Kartsonis N., Hanna G.J.,
RA      Wong J., Finzi D., Rosenberg E., Gunthard H.F., Sutton L., Savara A.,
RA      Petropoulos C.J., Hellmann N., Walker B.D., Richman D.D.,
RA      Siliciano R., D'Aquila R.T.;
RT      "Antiretroviral resistance during successful therapy of HIV type 1
RT      infection";
RL      Proc. Natl. Acad. Sci. U.S.A. 97:10948-10953(2000).
DR      EMBL; AF292890; AAG25539.1; -.
DR      HSSP; P05888; 1AAF.
DR      GO; GO:0019012; C:virion; IEA.
DR      GO; GO:0003676; F:nucleic acid binding; IEA.
DR      InterPro; IPR001878; Znf_CCHC.
DR      Pfam; PF00098; Zf-CCHC; 1.
DR      PRINTS; PR00939; C2HCZNFINGER.
DR      SMART; SM00343; Znf_CCHC; 1.
DR      PROSITE; PS50158; ZF_CCHC; 1.
KW      Core protein; Polyprotein.
FT      NON_TER      1
SQ      SEQUENCE      94 AA; 10659 MW; 42EABE775ED726C7 CRC64;

Query Match      84.1%; Score 37; DB 15; Length 94;
Best Local Similarity 75.0%; Pred. No. 31;
Matches      6; Conservative      2; Mismatches      0; Indels      0; Gaps      0;

QY      1 KPSPPEE 8
      :||:||||
Db      48 EPSAPPEE 55

RESULT 25
09DY07      PRELIMINARY;      PRT;      94 AA.
AC      09DY07;
DT      01-MAR-2001 (TEMBLrel. 16, Created)
DT      01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DE      01-OCT-2003 (TEMBLrel. 25, Last annotation update)
GN      Gag polyprotein (Fragment).
OS      Human immunodeficiency virus 1.
OC      Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX      NCBI_TaxID=11676;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=20461476; PubMed=11005867;
RA      Martinez-Picado J., DePasquale M.P., Kartsonis N., Hanna G.J.,
RA      Wong J., Finzi D., Rosenberg E., Gunthard H.F., Sutton L., Savara A.,
RA      Petropoulos C.J., Hellmann N., Walker B.D., Richman D.D.,
RA      Siliciano R., D'Aquila R.T.;
RT      "Antiretroviral resistance during successful therapy of HIV type 1
RT      infection";
RL      Proc. Natl. Acad. Sci. U.S.A. 97:10948-10953(2000).
DR      EMBL; AF292886; AAG25533.1; -.
DR      HSSP; P05888; 1AAF.

```

```

DR      GO; GO:0019012; C:virion; IEA.
DR      GO; GO:0003676; F:nucleic acid binding; IEA.
DR      InterPro; IPR001878; Znf_CCHC.
DR      Pfam; PF00098; Zf-CCHC; 1.
DR      PRINTS; PR00939; C2HCZNFINGER.
DR      SMART; SM00343; Znf_CCHC; 1.
DR      PROSITE; PS50158; ZF_CCHC; 1.
KW      Core protein; Polyprotein.
FT      NON_TER      1
SQ      SEQUENCE      94 AA; 10586 MW; 2997A0CE5ED73296 CRC64;

Query Match      84.1%; Score 37; DB 15; Length 94;
Best Local Similarity 75.0%; Pred. No. 31;
Matches      6; Conservative      2; Mismatches      0; Indels      0; Gaps      0;

QY      1 KPSPPEE 8
      :||:||||
Db      48 EPSAPPEE 55

RESULT 26
038202      PRELIMINARY;      PRT;      95 AA.
AC      038202;
DT      01-JAN-1998 (TEMBLrel. 05, Created)
DT      01-JAN-1998 (TEMBLrel. 05, Last sequence update)
DE      01-OCT-2003 (TEMBLrel. 25, Last annotation update)
GN      Gag protein (Fragment).
OS      Gag.
ON      Human immunodeficiency virus 1.
OC      Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX      NCBI_TaxID=11676;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97404676; PubMed=9261388;
RA      Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RA      Vasudevachari M.B., Salzman N.P.;
RT      "Drug resistance during indinavir therapy is caused by mutations in
RT      the protease gene and in its Gag substrate cleavage sites.";
RL      J. Virol. 71:6662-6670(1997).
DR      EMBL; AF024099; AAB82975.1; -.
DR      HSSP; P05888; 1AAF.
DR      GO; GO:0019012; C:virion; IEA.
DR      GO; GO:0003676; F:nucleic acid binding; IEA.
DR      InterPro; IPR001878; Znf_CCHC.
DR      Pfam; PF00098; Zf-CCHC; 2.
DR      PRINTS; PR00939; C2HCZNFINGER.
DR      SMART; SM00343; Znf_CCHC; 2.
DR      PROSITE; PS50158; ZF_CCHC; 1.
KW      Core protein; Polyprotein.
FT      NON_TER      1
FT      NON_TER      95
SQ      SEQUENCE      95 AA; 10631 MW; 01409EBE8FE6D26 CRC64;

Query Match      84.1%; Score 37; DB 15; Length 95;
Best Local Similarity 75.0%; Pred. No. 31;
Matches      6; Conservative      2; Mismatches      0; Indels      0; Gaps      0;

QY      1 KPSPPEE 8
      :||:||||
Db      61 EPSAPPEE 68

RESULT 27
038182      PRELIMINARY;      PRT;      95 AA.
AC      038182;
DT      01-JAN-1998 (TEMBLrel. 05, Created)
DT      01-JAN-1998 (TEMBLrel. 05, Last sequence update)
DE      01-OCT-2003 (TEMBLrel. 25, Last annotation update)
GN      Gag protein (Fragment).
OS      Human immunodeficiency virus 1.

```

OC viruses; Retrovird viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97404676; PubMed=9261388;  
 RA Zhang Y.M., Imanishi H., Imanishi T., Lane H.C., Falloon J.,  
 Vasudevachari M.B., Salzman N.P.;  
 RT "Drug resistance during indinavir therapy is caused by mutations in  
 the protease gene and in its Gag substrate cleavage sites.";  
 RL J. Virol. 71:6662-6670(1997).  
 DR EMBL; AF024079; AAB82955.1; -.  
 DR HSPSP; P05888; IAAF.  
 OS Human immunodeficiency virus 1.  
 DR GO; GO:0019012; C:viral; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00098; zf\_CCHC; 2.  
 DR PRINTS; PR00939; CZHCZNFINGER.  
 DR SMART; SM00343; Znf\_C2HC; 2.  
 DR PROSITE; PS50158; ZF\_CCHC; 1.  
 KW Core protein; Polyprotein.  
 FT NON TER 1  
 FT NON TER 95  
 SQ SEQUENCE 95 AA; 10690 MW; F1E21554E9B6D3A CRC64;

Query Match 84.1%; Score 37; DB 15; Length 95;  
 Best Local Similarity 75.0%; Pred. No. 31;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPE 8  
 Db 61 EPSAPPE 68

## RESULT 28

038195 PRELIMINARY; PRT; 96 AA.  
 AC 038195;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Gag protein (Fragment).  
 GN GAG.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retrovird viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97404676; PubMed=9261388;  
 RA Zhang Y.M., Imanishi H., Imanishi T., Lane H.C., Falloon J.,  
 Vasudevachari M.B., Salzman N.P.;  
 RT "Drug resistance during indinavir therapy is caused by mutations in  
 the protease gene and in its Gag substrate cleavage sites.";  
 RL J. Virol. 71:6662-6670(1997).  
 DR EMBL; AF024092; AAB82968.1; -.  
 DR HSPSP; P05888; IAAF.  
 OS Human immunodeficiency virus 1.  
 DR GO; GO:0019012; C:viral; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00098; zf\_CCHC; 2.  
 DR PRINTS; PR00939; CZHCZNFINGER.  
 DR SMART; SM00343; Znf\_C2HC; 2.  
 DR PROSITE; PS50158; ZF\_CCHC; 2.  
 KW Core protein; Polyprotein.  
 FT NON TER 1  
 FT NON TER 96  
 SQ SEQUENCE 96 AA; 10860 MW; C21589A930BCE78 CRC64;

Query Match 84.1%; Score 37; DB 15; Length 96;  
 Best Local Similarity 75.0%; Pred. No. 32;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPE 8  
 Db 61 EPSAPPE 68

Db 62 EPSAPPE 69

## RESULT 29

038214 PRELIMINARY; PRT; 96 AA.  
 AC 038214;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Gag protein (Fragment).  
 GN GAG.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retrovird viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97404676; PubMed=9261388;  
 RA Zhang Y.M., Imanishi H., Imanishi T., Lane H.C., Falloon J.,  
 Vasudevachari M.B., Salzman N.P.;  
 RT "Drug resistance during indinavir therapy is caused by mutations in  
 the protease gene and in its Gag substrate cleavage sites.";  
 RL J. Virol. 71:6662-6670(1997).  
 DR EMBL; AF024111; AAB82987.1; -.  
 DR HSPSP; P05888; IAAF.  
 OS Human immunodeficiency virus 1.  
 DR GO; GO:0019012; C:viral; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00098; zf\_CCHC; 2.  
 DR PRINTS; PR00939; CZHCZNFINGER.  
 DR PROSITE; PS50158; ZF\_CCHC; 2.  
 KW Core protein; Polyprotein.  
 FT NON TER 1  
 FT NON TER 96  
 SQ SEQUENCE 96 AA; 10838 MW; DBE103AA90B8CD78 CRC64;

Query Match 84.1%; Score 37; DB 15; Length 96;  
 Best Local Similarity 75.0%; Pred. No. 32;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPE 8  
 Db 62 EPSAPPE 69

## RESULT 30

038215 PRELIMINARY; PRT; 96 AA.  
 AC 038215;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Gag protein (Fragment).  
 GN GAG.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retrovird viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97404676; PubMed=9261388;  
 RA Zhang Y.M., Imanishi H., Imanishi T., Lane H.C., Falloon J.,  
 Vasudevachari M.B., Salzman N.P.;  
 RT "Drug resistance during indinavir therapy is caused by mutations in  
 the protease gene and in its Gag substrate cleavage sites.";  
 RL J. Virol. 71:6662-6670(1997).  
 DR EMBL; AF024112; AAB82988.1; -.  
 DR HSPSP; P05888; IAAF.  
 OS Human immunodeficiency virus 1.  
 DR GO; GO:0019012; C:viral; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00098; zf\_CCHC; 2.  
 DR PRINTS; PR00939; CZHCZNFINGER.  
 DR SMART; SM00343; Znf\_C2HC; 2.

```

DR PROSITE; PS50158; 2P CCHC; 2.
KW Core protein; Polyprotein.
FT NON_TER 1
SQ SEQUENCE 96 AA; 10910 MW; 79E303AA90B8CD78 CRC64;

Query Match
Best Local Similarity 84.1%; Score 37; DB 15; Length 96;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8
Db 62 EPSAPPPE 69

RESULT 31
038208 PRELIMINARY; PRT; 96 AA.
AC 038208;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Gag protein (Fragment).
GN GAG.
OS Human immunodeficiency virus 1.
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN (1)
RP MEDLINE=97404676; PubMed=9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024105; AAB82981.1; -.
DR HSSP; P05888; IAAF.
DR GO; GO:0019012; C:vifion; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001878; ZnF_CCHC.
DR Pfam; PF00098; zf-CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR PROSITE; PS50158; ZF CCHC; 2.
KW Core protein; Polyprotein.
FT NON_TER 1
FT NON_TER 96
SQ SEQUENCE 96 AA; 10838 MW; DBE103AA90B8CD78 CRC64;

Query Match
Best Local Similarity 84.1%; Score 37; DB 15; Length 96;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8
Db 62 EPSAPPPE 69

RESULT 32
038218 PRELIMINARY; PRT; 96 AA.
AC 038218;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Gag protein (Fragment).
GN GAG.
OS Human immunodeficiency virus 1.
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN (1)
RP MEDLINE=97404676; PubMed=9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024107; AAB82983.1; -.
DR HSSP; P05888; IAAF.
DR GO; GO:0019012; C:vifion; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001878; ZnF_CCHC.
DR Pfam; PF00098; zf-CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00343; ZnF_CCHC; 2.
DR PROSITE; PS50158; ZF CCHC; 2.
KW Core protein; Polyprotein.
FT NON_TER 1
FT NON_TER 96
SQ SEQUENCE 96 AA; 10817 MW; 330C03AA90A4C6FB CRC64;

Query Match
Best Local Similarity 84.1%; Score 37; DB 15; Length 96;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8
Db 62 EPSAPPPE 69

RESULT 33
038210 PRELIMINARY; PRT; 96 AA.
AC 038210;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Gag protein (Fragment).
GN GAG.
OS Human immunodeficiency virus 1.
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN (1)
RP MEDLINE=97404676; PubMed=9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024107; AAB82983.1; -.
DR HSSP; P05888; IAAF.
DR GO; GO:0019012; C:vifion; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001878; ZnF_CCHC.
DR Pfam; PF00098; zf-CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00343; ZnF_CCHC; 2.
DR PROSITE; PS50158; ZF CCHC; 2.
KW Core protein; Polyprotein.
FT NON_TER 1
FT NON_TER 96
SQ SEQUENCE 96 AA; 10817 MW; 330C03AA90A4C6FB CRC64;

Query Match
Best Local Similarity 84.1%; Score 37; DB 15; Length 96;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8
Db 62 EPSAPPPE 69

RESULT 34
038193 PRELIMINARY; PRT; 96 AA.
AC 038193;

```

```

RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024115; AAB82991.1; -.
DR HSSP; P05888; IAAF.
DR GO; GO:0019012; C:vifion; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001878; ZnF_CCHC.
DR Pfam; PF00098; zf-CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR PROSITE; PS50158; ZF CCHC; 2.
KW Core protein; Polyprotein.
FT NON_TER 1
FT NON_TER 96
SQ SEQUENCE 96 AA; 10838 MW; DBE103AA90B8CD78 CRC64;

Query Match
Best Local Similarity 84.1%; Score 37; DB 15; Length 96;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8
Db 62 EPSAPPPE 69

RESULT 33
038210 PRELIMINARY; PRT; 96 AA.
AC 038210;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Gag protein (Fragment).
GN GAG.
OS Human immunodeficiency virus 1.
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN (1)
RP MEDLINE=97404676; PubMed=9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024107; AAB82983.1; -.
DR HSSP; P05888; IAAF.
DR GO; GO:0019012; C:vifion; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001878; ZnF_CCHC.
DR Pfam; PF00098; zf-CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00343; ZnF_CCHC; 2.
DR PROSITE; PS50158; ZF CCHC; 2.
KW Core protein; Polyprotein.
FT NON_TER 1
FT NON_TER 96
SQ SEQUENCE 96 AA; 10817 MW; 330C03AA90A4C6FB CRC64;

Query Match
Best Local Similarity 84.1%; Score 37; DB 15; Length 96;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8
Db 62 EPSAPPPE 69

RESULT 34
038193 PRELIMINARY; PRT; 96 AA.
AC 038193;

```

```

DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Gag protein (Fragment).
GN
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN
RP MEDLINE=97404676; PubMed=9261388;
RX Zhang Y.M., Imanichi H., Imanichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR HSP; P05888; 1AA.
DR GO; GO:0019012; C:viral; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf_CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00343; Znf_CCHC; 2.
DR PROSITE; PS50158; ZF_CCHC; 2.
DR Core protein; Polyprotein.
FT NON TER 1
FT 96
SQ SEQUENCE 96 AA; 10893 MW; 5A708AAA90A82376 CRC64;

Query Match 84.1%; Score 37; DB 15; Length 96;
Best Local Similarity 75.0%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPE 8
Db 62 EPSAPPE 69

RESULT 35
ID 038183 PRELIMINARY; PRT; 96 AA.
AC 038183;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Gag protein (Fragment).
GN
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN
RP MEDLINE=97404676; PubMed=9261388;
RX Zhang Y.M., Imanichi H., Imanichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR HSP; P05888; 1AA.
DR GO; GO:0019012; C:viral; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf_CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00343; Znf_CCHC; 2.
DR PROSITE; PS50158; ZF_CCHC; 2.
DR Core protein; Polyprotein.
FT NON TER 1
FT 96
SQ SEQUENCE 96 AA; 10778 MW; 328FA21B20B8CD64 CRC64;

```

```

Query Match 84.1%; Score 37; DB 15; Length 96;
Best Local Similarity 75.0%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPE 8
Db 62 EPSAPPE 69

RESULT 36
ID 038204 PRELIMINARY; PRT; 96 AA.
AC 038204;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Gag protein (Fragment).
GN
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN
RP MEDLINE=97404676; PubMed=9261388;
RX Zhang Y.M., Imanichi H., Imanichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR HSP; P05888; 1AA.
DR GO; GO:0019012; C:viral; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf_CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR PROSITE; PS50158; ZF_CCHC; 2.
DR Core protein; Polyprotein.
FT NON TER 1
FT 96
SQ SEQUENCE 96 AA; 10893 MW; 5A708AAA90A82376 CRC64;

Query Match 84.1%; Score 37; DB 15; Length 96;
Best Local Similarity 75.0%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPE 8
Db 62 EPSAPPE 69

RESULT 37
ID 038221 PRELIMINARY; PRT; 96 AA.
AC 038221;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Gag protein (Fragment).
GN
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN
RP MEDLINE=97404676; PubMed=9261388;
RX Zhang Y.M., Imanichi H., Imanichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR HSP; P05888; 1AA.

```

```

DR GO:0019012; C.vitron; IEA.
DR GO:0003676; F.nucleic acid binding; IEA.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00098; Zf-CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00343; Znf_CZHC; 2.
DR PROSITE; PS50158; ZF_CCHC; 2.
KW Core protein; Polypeptide.
FT NON_TER 1 1
FT NON_TER 96 96
SQ SEQUENCE 96 AA; 10894 MW; D46A82B208B15CF1 CRC64;

Query Match
Best Local Similarity 84.1%; Score 37; DB 15; Length 96;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8
Db 62 EPSAPPPEE 69

RESULT 38
O38205 PRELIMINARY; PRT; 96 AA.
AC O38205;
DT 01-JAN-1998 (TEMBLrel. 05, Created)
DT 01-JUN-1998 (TEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Gag protein (Fragment).
GN GAG.
OS Human immunodeficiency virus 1.
OC Viruses; Retrovirdae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97404676; PubMed=9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024102; AAB82978.1; -.
DR HSSP; P05888; IAAF.
DR GO:0019012; C.vitron; IEA.
DR GO:0003676; F.nucleic acid binding; IEA.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam; PF00098; Zf-CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR PROSITE; PS50158; ZF_CCHC; 2.
KW Core protein; Polypeptide.
FT NON_TER 1 1
FT NON_TER 96 96
SQ SEQUENCE 96 AA; 10778 MW; 328FA21B20B8CD64 CRC64;

Query Match
Best Local Similarity 84.1%; Score 37; DB 15; Length 96;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8
Db 62 EPSAPPPEE 69

RESULT 39
O38206 PRELIMINARY; PRT; 96 AA.
AC O38206;
DT 01-JAN-1998 (TEMBLrel. 05, Created)
DT 01-JUN-1998 (TEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Gag protein (Fragment).
GN GAG.
OS Human immunodeficiency virus 1.

```

```

OC Viruses; Retrovirdae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97404676; PubMed=9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024103; AAB82979.1; -.
DR HSSP; P05888; IAAF.
DR GO:0019012; C.vitron; IEA.
DR GO:0003676; F.nucleic acid binding; IEA.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam; PF00098; Zf-CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00343; Znf_CZHC; 2.
DR PROSITE; PS50158; ZF_CCHC; 2.
KW Core protein; Polypeptide.
FT NON_TER 1 1
FT NON_TER 96 96
SQ SEQUENCE 96 AA; 10838 MW; DBE103AA90B8CD78 CRC64;

Query Match
Best Local Similarity 84.1%; Score 37; DB 15; Length 96;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8
Db 62 EPSAPPPEE 69

RESULT 40
O38198 PRELIMINARY; PRT; 96 AA.
AC O38198;
DT 01-JAN-1998 (TEMBLrel. 05, Created)
DT 01-JUN-1998 (TEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Gag protein (Fragment).
GN GAG.
OS Human immunodeficiency virus 1.
OC Viruses; Retrovirdae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97404676; PubMed=9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024095; AAB82971.1; -.
DR HSSP; P05888; IAAF.
DR GO:0019012; C.vitron; IEA.
DR GO:0003676; F.nucleic acid binding; IEA.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam; PF00098; Zf-CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR PROSITE; PS50158; ZF_CCHC; 2.
KW Core protein; Polypeptide.
FT NON_TER 1 1
FT NON_TER 96 96
SQ SEQUENCE 96 AA; 10810 MW; F6C269BE6748CD66 CRC64;

Query Match
Best Local Similarity 84.1%; Score 37; DB 15; Length 96;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPEE 8
Db 62 EPSAPPPEE 69

```

```

RESULT 41
038207 ID 038207 PRELIMINARY; PRT; 96 AA.
AC 038207;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Gag protein (Fragment).
GN GAG.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP MEDLINE=97404676; PubMed=9261388;
RA Zhang Y.M., Imanishi H., Imanishi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL: AF024104; AAB82980.1; -.
DR HSP: P05888; 1A8F.
DR GO: GO:0019012; C:viral; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00098; zf-CCHC; 2.
DR PRINTS: PR00939; C2HCZNFINGER.
DR SMART: SM00343; Znf_CCHC; 2.
DR PROSITE: PS50158; ZF_CCHC; 2.
DR Core protein; Polyprotein.
FT NON TER 1
FT NON TER 96
SQ SEQUENCE 96 AA; 10778 MW; FD3A7FCA90B8CD65 CRC64;

Query Match 84.1%; Score 37; DB 15; Length 96;
Best Local Similarity 75.0%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPEE 8
Db 62 EPSAPPE 69

RESULT 42
038197 ID 038197 PRELIMINARY; PRT; 96 AA.
AC 038197;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Gag protein (Fragment).
GN GAG.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP MEDLINE=97404676; PubMed=9261388;
RA Zhang Y.M., Imanishi H., Imanishi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL: AF024094; AAB82970.1; -.
DR HSP: P05888; 1A8F.
DR GO: GO:0019012; C:viral; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00098; zf-CCHC; 2.
DR PRINTS: PR00939; C2HCZNFINGER.
DR SMART: SM00343; Znf_CCHC; 2.
DR PROSITE: PS50158; ZF_CCHC; 1.
DR Core protein; Polyprotein.
FT NON TER 1
FT NON TER 96
SQ SEQUENCE 96 AA; 10778 MW; FD3A7FCA90B8CD65 CRC64;

Query Match 84.1%; Score 37; DB 15; Length 96;
Best Local Similarity 75.0%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPEE 8
Db 62 EPSAPPE 69

```

```

DR PROSITE; PS50158; ZF_CCHC; 2.
KW Core protein; Polyprotein.
FT NON TER 1
FT NON TER 96
SQ SEQUENCE 96 AA; 10810 MW; F6C269BE6748CD66 CRC64;

Query Match 84.1%; Score 37; DB 15; Length 96;
Best Local Similarity 75.0%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPEE 8
Db 62 EPSAPPE 69

RESULT 43
038184 ID 038184 PRELIMINARY; PRT; 96 AA.
AC 038184;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Gag protein (Fragment).
GN GAG.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP MEDLINE=97404676; PubMed=9261388;
RA Zhang Y.M., Imanishi H., Imanishi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL: AF024081; AAB82957.1; -.
DR HSP: P05888; 1A8F.
DR GO: GO:0019012; C:viral; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00098; zf-CCHC; 1.
DR PRINTS: PR00939; C2HCZNFINGER.
DR SMART: SM00343; Znf_CCHC; 2.
DR PROSITE: PS50158; ZF_CCHC; 1.
DR Core protein; Polyprotein.
FT NON TER 1
FT NON TER 96
SQ SEQUENCE 96 AA; 10861 MW; 5F4E7B1AE0AD9971 CRC64;

Query Match 84.1%; Score 37; DB 15; Length 96;
Best Local Similarity 75.0%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPEE 8
Db 62 EPSAPPE 69

RESULT 44
038194 ID 038194 PRELIMINARY; PRT; 96 AA.
AC 038194;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Gag protein (Fragment).
GN GAG.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP MEDLINE=97404676; PubMed=9261388;

```

RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,  
 RA Vasudevachari M.B., Salzman N.P.;  
 RT "Drug resistance during indinavir therapy is caused by mutations in  
 the protease gene and in its Gag substrate cleavage sites.";  
 RL J. Virol. 71:6662-6670(1997).  
 DR EMBL; AF024091; AAB82967.1; -.  
 DR HSSP; P05888; IAAF.  
 DR GO; GO:0019012; C.vitron; IEA.  
 DR GO; GO:0003676; F.nucleic acid binding; IEA.  
 DR InterPro; IPR001878; ZnF\_CCHC.  
 DR Pfam; PF00098; zf-CCHC; 2.  
 DR PRINTS; PR00939; C2HCZNFINGER.  
 DR PROSITE; PS50158; ZF\_CCHC; 2.  
 KW Core protein; Polyprotein.  
 FT NON\_TER 1  
 FT SEQUENCE 96 AA; 10778 MW; 328FA21B20B8CD64 CRC64;

Query Match 84.1%; Score 37; DB 15; Length 96;  
 Best Local Similarity 75.0%; Pred. No. 32;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPBE 8  
 Db 62 EPSAPBE 69

RESULT 45  
 ID 038200 PRELIMINARY; PRT; 96 AA.

AC 038200;  
 DT 01-JAN-1998 (TEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)  
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
 DE Gag protein (Fragment).  
 GN GAG.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OC NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97404676; PubMed=9261388;  
 RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,  
 RA Vasudevachari M.B., Salzman N.P.;  
 RT "Drug resistance during indinavir therapy is caused by mutations in  
 the protease gene and in its Gag substrate cleavage sites.";  
 RL J. Virol. 71:6662-6670(1997).  
 DR EMBL; AF024097; AAB82973.1; -.  
 DR HSSP; P05888; IAAF.  
 DR GO; GO:0019012; C.vitron; IEA.  
 DR GO; GO:0003676; F.nucleic acid binding; IEA.  
 DR InterPro; IPR001878; ZnF\_CCHC.  
 DR Pfam; PF00098; zf-CCHC; 2.  
 DR PRINTS; PR00939; C2HCZNFINGER.  
 DR SMART; SM00343; ZnF\_CCHC; 2.  
 DR PROSITE; PS50158; ZF\_CCHC; 2.  
 KW Core protein; Polyprotein.  
 FT NON\_TER 1  
 FT SEQUENCE 96 AA; 10790 MW; 32958B1B20B8CD64 CRC64;

Query Match 84.1%; Score 37; DB 15; Length 96;  
 Best Local Similarity 75.0%; Pred. No. 32;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPBE 8  
 Db 62 EPSAPBE 69

RESULT 46  
 ID 038211 PRELIMINARY; PRT; 96 AA.

AC 038211;  
 DT 01-JAN-1998 (TEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)  
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
 DE Gag protein (Fragment).  
 GN GAG.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OC NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97404676; PubMed=9261388;  
 RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,  
 RA Vasudevachari M.B., Salzman N.P.;  
 RT "Drug resistance during indinavir therapy is caused by mutations in  
 the protease gene and in its Gag substrate cleavage sites.";  
 RL J. Virol. 71:6662-6670(1997).  
 DR EMBL; AF024108; AAB82984.1; -.  
 DR HSSP; P05888; IAAF.  
 DR GO; GO:0019012; C.vitron; IEA.  
 DR GO; GO:0003676; F.nucleic acid binding; IEA.  
 DR InterPro; IPR001878; ZnF\_CCHC.  
 DR Pfam; PF00098; zf-CCHC; 2.  
 DR PRINTS; PR00939; C2HCZNFINGER.  
 DR PROSITE; PS50158; ZF\_CCHC; 2.  
 KW Core protein; Polyprotein.  
 FT NON\_TER 1  
 FT SEQUENCE 96 AA; 10838 MW; DBE103AA90B8CD78 CRC64;

Query Match 84.1%; Score 37; DB 15; Length 96;  
 Best Local Similarity 75.0%; Pred. No. 32;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPBE 8  
 Db 62 EPSAPBE 69

RESULT 47  
 ID 038181 PRELIMINARY; PRT; 96 AA.

AC 038181;  
 DT 01-JAN-1998 (TEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)  
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
 DE Gag protein (Fragment).  
 GN GAG.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OC NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97404676; PubMed=9261388;  
 RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,  
 RA Vasudevachari M.B., Salzman N.P.;  
 RT "Drug resistance during indinavir therapy is caused by mutations in  
 the protease gene and in its Gag substrate cleavage sites.";  
 RL J. Virol. 71:6662-6670(1997).  
 DR EMBL; AF024078; AAB82954.1; -.  
 DR HSSP; P05888; IAAF.  
 DR GO; GO:0019012; C.vitron; IEA.  
 DR GO; GO:0003676; F.nucleic acid binding; IEA.  
 DR InterPro; IPR001878; ZnF\_CCHC.  
 DR Pfam; PF00098; zf-CCHC; 2.  
 DR PRINTS; PR00939; C2HCZNFINGER.  
 DR SMART; SM00343; ZnF\_CCHC; 2.  
 DR PROSITE; PS50158; ZF\_CCHC; 2.  
 KW Core protein; Polyprotein.  
 FT NON\_TER 1  
 FT SEQUENCE 96 AA; 10810 MW; BEA6E10FD748CD67 CRC64;

Query Match 84.1%; Score 37; DB 15; Length 96;  
 Best Local Similarity 75.0%; Pred. No. 32;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
 DB 62 EPSAPPE 69

## RESULT 48

OC 038216 PRELIMINARY; PRT; 96 AA.  
 AC 038216;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Gag protein (Fragment).  
 GN GAG.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=97404676; PubMed=9261388;  
 RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,  
 RT "Drug resistance during indinavir therapy is caused by mutations in  
 the protease gene and in its Gag substrate cleavage sites.";  
 RL J. Virol. 71:6662-6670(1997).  
 DR EMBL; AF024113; AAB82989.1; -.  
 DR HSSP; P05888; IAAF.

DR GO; GO:0019012; C:Viron; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR InterPro: IPR001878; Znf CCHC.  
 DR Pfam: PF00098; zf-CCHC; 2.  
 DR PRINTS; PR00939; C2HCZFRINGER.  
 DR PROSITE; PS50158; ZF\_CCHC; 2.  
 KW Core protein; Polyprotein.  
 FT NON TER 1  
 FT NON TER 96  
 SQ SEQUENCE 96 AA; 10838 MW; DBE103AA90B8CD78 CRC64;

Query Match 84.1%; Score 37; DB 15; Length 96;  
 Best Local Similarity 75.0%; Pred. No. 32;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
 DB 62 EPSAPPE 69

## RESULT 49

OC 038213 PRELIMINARY; PRT; 96 AA.  
 AC 038213;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Gag protein (Fragment).  
 GN GAG.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=97404676; PubMed=9261388;  
 RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,  
 RT "Drug resistance during indinavir therapy is caused by mutations in  
 the protease gene and in its Gag substrate cleavage sites.";  
 RL J. Virol. 71:6662-6670(1997).  
 DR EMBL; AF024110; AAB82986.1; -.  
 DR HSSP; P05888; IAAF.

DR GO; GO:0019012; C:Viron; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR InterPro: IPR001878; Znf CCHC.  
 DR Pfam: PF00098; zf-CCHC; 2.  
 DR PRINTS; PR00939; C2HCZFRINGER.  
 DR PROSITE; PS50158; ZF\_CCHC; 2.  
 KW Core protein; Polyprotein.  
 FT NON TER 1  
 FT NON TER 96  
 SQ SEQUENCE 96 AA; 10838 MW; DBE103AA90B8CD78 CRC64;

Query Match 84.1%; Score 37; DB 15; Length 96;  
 Best Local Similarity 75.0%; Pred. No. 32;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
 DB 62 EPSAPPE 69

## RESULT 50

OC 038222 PRELIMINARY; PRT; 96 AA.  
 AC 038222;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Gag protein (Fragment).  
 GN GAG.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=97404676; PubMed=9261388;  
 RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,  
 RT "Drug resistance during indinavir therapy is caused by mutations in  
 the protease gene and in its Gag substrate cleavage sites.";  
 RL J. Virol. 71:6662-6670(1997).  
 DR EMBL; AF024119; AAB82995.1; -.  
 DR HSSP; P05888; IAAF.

DR GO; GO:0019012; C:Viron; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR InterPro: IPR001878; Znf CCHC.  
 DR Pfam: PF00098; zf-CCHC; 2.  
 DR PRINTS; PR00939; C2HCZFRINGER.  
 DR SMART; SM00343; ZNF\_CCHC; 2.  
 DR PROSITE; PS50158; ZF\_CCHC; 2.  
 KW Core protein; Polyprotein.  
 FT NON TER 1  
 FT NON TER 96  
 SQ SEQUENCE 96 AA; 10852 MW; C3EB2AA81AD9C78 CRC64;

Query Match 84.1%; Score 37; DB 15; Length 96;  
 Best Local Similarity 75.0%; Pred. No. 32;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
 DB 62 EPSAPPE 69

Search completed: June 18, 2004, 12:59:58  
 Job time : 52 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2004, 12:56:24 ; Search time 22 Seconds  
(without alignments)  
34.979 Million cell updates/sec

Title: US-10-655-201-2

Perfect score: 44

Sequence: 1 KPSPPEE 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 500 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	100.0	431	1 UKHU	u-plasminogen acti
2	44	100.0	433	1 UKBAV	u-plasminogen acti
3	39	88.6	990	2 A86215	protein T6D22.8 [l
4	36	81.8	124	2 H87616	cytochrome c famil
5	36	81.8	261	2 C87204	thioesterase [impo
6	36	81.8	418	2 A47281	pigment epithelial
7	36	81.8	880	2 T02245	hypothetical prote
8	36	81.8	1185	2 T46428	hypothetical prote
9	36	81.8	1358	2 A29360	SIR4 protein - yea
10	36	81.8	1806	1 CGHUIE	collagen alpha 1(X
11	35	79.5	90	2 B83542	hypothetical prote
12	35	79.5	216	2 T26841	hypothetical prote
13	35	79.5	249	2 H72858	apoptosis inhibito
14	35	79.5	249	2 T41814	IAP2 orf171 - Bomby
15	35	79.5	374	2 T07887	G box-binding proc
16	35	79.5	376	2 T07882	mRNA 3'-end proces
17	35	79.5	407	2 D69316	protein-tyrosine-p
18	35	79.5	440	2 T50213	hypothetical prote
19	35	79.5	516	2 E96583	hypothetical prote
20	35	79.5	754	1 BABOH	peptide-aspartate
21	35	79.5	765	2 A53047	6-phosphofructokin
22	35	79.5	784	2 JC2055	6-phosphofructokin
23	35	79.5	791	2 A53206	6-phosphofructokin
24	35	79.5	1179	2 G75459	DNA-directed RNA p
25	35	79.5	1186	2 T12737	tail protein - Met
26	35	79.5	1240	1 DJBE21	DNA-directed DNA p
27	35	79.5	1691	1 D54689	protein-tyrosine-p
28	35	79.5	1863	2 S46217	protein-tyrosine-p
29	35	79.5	1894	2 C54689	protein-tyrosine-p

30	35	79.5	1907	2 S50893	protein-tyrosine-p
31	35	79.5	1912	2 A56178	protein-tyrosine-p
32	35	79.5	2265	1 FNBO	fibronectin - bovi
33	35	79.5	2386	1 FNBU	fibronectin precu
34	35	79.5	2477	2 S14428	fibronectin precu
35	34	77.3	66	2 AG2522	hypothetical prote
36	34	77.3	122	2 F83167	hypothetical prote
37	34	77.3	124	2 H82468	hypothetical prote
38	34	77.3	136	2 E72759	hypothetical prote
39	34	77.3	140	2 E72505	hypothetical prote
40	34	77.3	173	2 T31243	hypothetical prote
41	34	77.3	231	2 H82155	hypothetical prote
42	34	77.3	301	2 A60548	synaptophysin - Pa
43	34	77.3	310	2 T33972	hypothetical prote
44	34	77.3	382	2 G84900	G-box binding bzp
45	34	77.3	458	2 UC7868	triacylglycerol 11
46	34	77.3	462	2 A42401	macrophage elastas
47	34	77.3	500	1 FOVMIV	gag polyprotein -
48	34	77.3	500	1 FOVMH4	gag polyprotein -
49	34	77.3	500	2 S33979	gag polyprotein -
50	34	77.3	502	1 FOVMH2	gag polyprotein -
51	34	77.3	506	1 A38068	gag polyprotein -
52	34	77.3	512	1 FOVMH3	gag polyprotein -
53	34	77.3	544	2 S75109	cytochrome-c oxida
54	34	77.3	553	1 A42499	mullerian inhibiti
55	34	77.3	555	1 S20100	mullerian inhibiti
56	34	77.3	558	2 A46225	nuclear orphan rec
57	34	77.3	598	2 A57040	T-cell nuclear rec
58	34	77.3	598	2 I84692	nuclear orphan rec
59	34	77.3	615	2 T41576	hypothetical prote
60	34	77.3	711	2 T04455	hypothetical prote
61	34	77.3	760	2 F86387	probable pto kinas
62	34	77.3	795	2 T49835	hypothetical prote
63	34	77.3	825	2 F75508	conserved hypothec
64	34	77.3	826	2 H84683	hypothetical prote
65	34	77.3	866	2 A54442	3',5'-cyclic-nucle
66	34	77.3	928	2 S64350	hypothetical prote
67	34	77.3	1071	2 E85343	hypothetical prote
68	34	77.3	1488	2 T02856	probable membrane
69	34	77.3	1571	2 T00062	hypothetical prote
70	34	77.3	2023	2 T13154	polycomb protein e
71	34	77.3	3938	2 T42761	Bassoon protein -
72	34	77.3	3942	2 T42730	Bassoon protein -
73	34	77.3	7962	2 I38346	elastic titin - hu
74	33	75.0	133	2 B70338	general secretion
75	33	75.0	137	2 H70547	hypothetical prote
76	33	75.0	153	2 E47021	pectic enzyme secr
77	33	75.0	176	2 F72651	hypothetical prote
78	33	75.0	245	2 AD2269	hypothetical prote
79	33	75.0	301	2 C64633	hook assembly prot
80	33	75.0	326	2 PC4028	dihydrolipomide S
81	33	75.0	353	2 F71881	probable flagellar
82	33	75.0	366	2 A53286	cell-surface glyco
83	33	75.0	378	1 H69004	2-oxoglutarate-fer
84	33	75.0	393	2 S44202	acyl-lacyl-carrier
85	33	75.0	394	2 C84905	probable extensin
86	33	75.0	399	1 OHSPAD	acyl-lacyl-carrier
87	33	75.0	411	1 E71178	translation initia
88	33	75.0	411	2 F75163	translation initia
89	33	75.0	424	1 H69323	translation initia
90	33	75.0	445	2 AF0033	H+-transporting tw
91	33	75.0	506	1 FOLJG3	gag polyprotein -
92	33	75.0	514	2 A31643	cell adhesion 80K
93	33	75.0	514	2 A44100	cell adhesion mole
94	33	75.0	537	2 T24945	hypothetical prote
95	33	75.0	596	2 G87659	peptidase M3 famil
96	33	75.0	640	1 UZADP7	terminal protein p
97	33	75.0	830	1 B48723	replication licens
98	33	75.0	847	2 F96531	hypothetical prote
99	33	75.0	927	2 B82075	sensor histidine k
100	33	75.0	975	2 T08606	protein phosphatas
101	33	75.0	1014	1 S75023	sensory transducti
102	33	75.0	1264	2 A36858	G2R protein - vari

103	33	75.0	1296	1	HMSOIF	aggregation protei	176	31	70.5	133	2	AE3330	hypothetical prote
104	33	75.0	1305	2	H41662	150K mating aggreg	177	31	70.5	135	2	A72721	hypothetical prote
105	33	75.0	1306	2	S22624	aggregation protei	178	31	70.5	137	2	A75292	hypothetical prote
106	33	75.0	1896	2	B72175	D15R protein - var	179	31	70.5	156	2	D75605	hypothetical prote
107	33	75.0	1897	2	T28621	hypothetical prote	180	31	70.5	157	2	S19735	lectin precursor -
108	33	75.0	4957	2	T03455	ALK protein - huma	181	31	70.5	158	2	UC1095	pre-S protein - du
109	33	75.0	5262	2	T03454	ALK protein - huma	182	31	70.5	160	2	UE0136	lectin precursor -
110	32	72.7	101	2	T36269	probable redoxin -	183	31	70.5	165	2	G84767	glycine decarboxyl
111	32	72.7	115	2	A23925	proline-rich phosph	184	31	70.5	166	2	A86450	probable glycine c
112	32	72.7	138	2	A95913	hypothetical membr	185	31	70.5	179	2	T19557	hypothetical prote
113	32	72.7	170	2	F87236	conserved hypochet	186	31	70.5	187	2	S74987	hypothetical prote
114	32	72.7	174	1	RDSPTA	ferredoxin-choroet	187	31	70.5	187	2	E75591	hypothetical prote
115	32	72.7	178	2	E84049	hypothetical prote	188	31	70.5	196	2	T27097	hypothetical prote
116	32	72.7	207	2	A56190	tilin - rat (fragm	189	31	70.5	209	2	T20975	hypothetical prote
117	32	72.7	238	2	T40820	proline-rich prote	190	31	70.5	215	2	T27098	hypothetical prote
118	32	72.7	243	2	A31231	high-affinity IGE	191	31	70.5	232	2	T26209	hypothetical prote
119	32	72.7	247	2	H69030	coenzyme PQQ synth	192	31	70.5	226	2	C96794	hypothetical prote
120	32	72.7	251	2	D31957	tropomycin T, skelet	193	31	70.5	228	1	S35689	venomabin A (EC 3.4
121	32	72.7	253	2	B34327	tropomycin T, fast s	194	31	70.5	229	2	C96013	probable transcript
122	32	72.7	257	2	B31957	tropomycin T, skelet	195	31	70.5	237	2	G65084	hypothetical prote
123	32	72.7	263	2	C31957	tropomycin T, skelet	196	31	70.5	242	2	T27999	hypothetical prote
124	32	72.7	283	2	S13383	hydroxyproline-ric	197	31	70.5	250	2	T03595	L-ascorbate peroxi
125	32	72.7	303	2	S28264	hydroxyproline-ric	198	31	70.5	253	2	F86184	hypothetical prote
126	32	72.7	328	2	UQ0965	hydroxyproline-ric	199	31	70.5	268	2	H84684	En/Spm-like transp
127	32	72.7	332	2	E86448	hydroxyproline-ric	200	31	70.5	269	2	D64668	hypothetical prote
128	32	72.7	350	2	S22456	hydroxyproline-ric	201	31	70.5	277	2	I38857	microtubule-associ
129	32	72.7	369	2	S20500	hydroxyproline-ric	202	31	70.5	280	2	B87305	metallo-beta-lacta
130	32	72.7	384	2	AG3013	conserved hypochet	203	31	70.5	283	2	T51091	hypothetical prote
131	32	72.7	384	2	H98270	hypothetical prote	204	31	70.5	304	2	T44637	dipeptide transpor
132	32	72.7	405	2	T17271	hypothetical prote	205	31	70.5	308	2	I48080	coatomer complex e
133	32	72.7	421	2	S38948	serine-ENNA ligase	206	31	70.5	308	2	I46019	coatomer complex e
134	32	72.7	430	2	AD0138	ToIB colicin import	207	31	70.5	310	2	E84612	homeodomain transc
135	32	72.7	434	2	C71434	probable phosphat	208	31	70.5	311	2	S47136	hypothetical prote
136	32	72.7	436	2	I49714	MHC H-2Kf-W5-link	209	31	70.5	312	2	T19675	hypothetical prote
137	32	72.7	517	2	E89530	protein H28G03.2 [	210	31	70.5	314	2	T38955	hypothetical prote
138	32	72.7	532	2	TE0081	testis sodium chan	211	31	70.5	314	2	B70013	hypothetical prote
139	32	72.7	539	2	H71280	hypothetical prote	212	31	70.5	316	2	UC6549	apolipoprotein E p
140	32	72.7	549	1	TEBEHC	legument protein -	213	31	70.5	316	2	S26478	apolipoprotein E -
141	32	72.7	556	2	T03114	hypothetical protei	214	31	70.5	316	2	I45996	conserved hypochet
142	32	72.7	556	2	T47552	hypothetical prote	215	31	70.5	317	2	H75297	hypothetical prote
143	32	72.7	562	2	F72771	probable lysyl-tRN	216	31	70.5	318	2	S76295	hypothetical prote
144	32	72.7	592	2	T32402	hypothetical prote	217	31	70.5	322	2	E84105	oligopeptide ABC t
145	32	72.7	613	2	T42671	hypothetical prote	218	31	70.5	323	2	T29032	hypothetical prote
146	32	72.7	615	2	T25245	hypothetical prote	219	31	70.5	329	1	G3M5C	Ig gamma-3 chain C
147	32	72.7	632	2	AF3095	nitric oxide reduc	220	31	70.5	329	2	T10064	cytokinin-induced
148	32	72.7	632	2	D98191	trkA-like protein	221	31	70.5	332	2	D86295	hypothetical prote
149	32	72.7	685	2	T18964	hypothetical prote	222	31	70.5	339	2	T26328	hypothetical prote
150	32	72.7	741	2	A83271	hypothetical prote	223	31	70.5	344	1	SAVUD	large surface anti
151	32	72.7	783	2	T38891	hypothetical prote	224	31	70.5	359	2	AB2733	NADH ubiquinone ox
152	32	72.7	832	2	H72278	alpha-mannosidase-	225	31	70.5	359	2	D97514	complex I 24k chat
153	32	72.7	839	2	F75518	hypothetical prote	226	31	70.5	370	2	T33382	hypothetical prote
154	32	72.7	875	2	T50182	ubiquitin-specific	227	31	70.5	372	2	T31060	hypothetical prote
155	32	72.7	950	2	A82986	adenylate cyclase	228	31	70.5	382	2	A88099	protein F18A12.2 [
156	32	72.7	953	2	T01093	luminidependens pr	229	31	70.5	387	2	T10793	acyl-lacyl-carrier
157	32	72.7	988	2	T25541	hypothetical prote	230	31	70.5	398	1	G3M5M	Ig gamma-3 chain C
158	32	72.7	1234	2	T30254	jumonji protein -	231	31	70.5	398	2	S23351	acyl-lacyl-carrier
159	32	72.7	1249	2	H71404	hypothetical prote	232	31	70.5	405	2	T42992	cyclin D - Caenorh
160	32	72.7	1294	2	T48349	EN2 protein - Ara	233	31	70.5	405	2	T26678	hypothetical prote
161	32	72.7	2054	2	T46612	multi-PDZ domain p	234	31	70.5	407	2	B47757	retrovirus-related
162	32	72.7	2701	2	S17796	inositol-trisphosp	235	31	70.5	415	2	S29345	translational elonga
163	31	70.5	66	2	AB2510	hypothetical prote	236	31	70.5	416	2	S48957	hypothetical prote
164	31	70.5	74	2	T47211	ccg-6 protein (imp	237	31	70.5	418	2	F71460	probable membrane
165	31	70.5	77	2	S30873	hypothetical prote	238	31	70.5	444	2	F96838	hypothetical prote
166	31	70.5	78	2	H69790	conserved hypochet	239	31	70.5	445	2	B40970	undulin 2 - human
167	31	70.5	81	2	S35686	high potential iro	240	31	70.5	446	2	A34418	H-2 region II bind
168	31	70.5	83	1	IHKREG	high potential iro	241	31	70.5	448	2	D41727	retinoid X recepto
169	31	70.5	85	1	IHKREV	high potential iro	242	31	70.5	458	2	T16041	hypothetical prote
170	31	70.5	94	2	A01955	Ig kappa-B5 chain	243	31	70.5	468	2	H70427	replicative DNA he
171	31	70.5	97	2	S11755	hypothetical prote	244	31	70.5	474	2	E87650	peptidase, M20/M25
172	31	70.5	99	2	T33486	hypothetical prote	245	31	70.5	476	2	A70477	glutamate synthase
173	31	70.5	101	2	G72155	R7R protein - var	246	31	70.5	476	1	S57963	methyl CPG binding
174	31	70.5	101	2	C36841	R7R protein - var	247	31	70.5	480	1	DEGRA	aromatic-L-amino-a
175	31	70.5	101	2	T28479	hypothetical prote	248	31	70.5	487	2	E70961	hypothetical prote

249	31	70.5	496	2	A41264	glucose transport
250	31	70.5	503	2	T10944	cysteine proteinase
251	31	70.5	511	2	C56849	dopamine receptor-
252	31	70.5	518	2	G86454	CDS protein F9L11.
253	31	70.5	520	2	T184718	RXR-beta1 isoform
254	31	70.5	527	2	A32469	80K protein H prec
255	31	70.5	534	2	JC7912	amino-acid N-acetyl
256	31	70.5	536	2	B84549	probable ubiquitin
257	31	70.5	538	2	T16034	hypothetical prote
258	31	70.5	540	2	AC0063	hypothetical prote
259	31	70.5	551	2	G95176	conserved hypothet
260	31	70.5	551	2	E84106	hypothetical prote
261	31	70.5	551	2	A98043	hypothetical prote
262	31	70.5	576	2	AC2195	hypothetical prote
263	31	70.5	579	2	T24536	hypothetical prote
264	31	70.5	580	2	B70668	probable Acyl-CoA
265	31	70.5	585	2	E70930	hypothetical prote
266	31	70.5	588	2	B70618	probable PE protei
267	31	70.5	601	2	T37738	hypothetical prote
268	31	70.5	611	1	S12566	translation initia
269	31	70.5	637	2	AB2403	ABC transporter AT
270	31	70.5	639	2	DB3591	probable ATP-depen
271	31	70.5	656	1	A34890	histidine decarbox
272	31	70.5	657	2	H71422	hypothetical prote
273	31	70.5	679	2	H84516	hypothetical prote
274	31	70.5	682	1	S22700	amphiphysin - chic
275	31	70.5	684	2	T37944	hypothetical prote
276	31	70.5	687	1	PYFFW	white protein - fr
277	31	70.5	695	2	S62400	amphiphysin (clone
278	31	70.5	731	2	T09172	probable calcium-a
279	31	70.5	748	2	T49633	glucan 1,4-alpha-g
280	31	70.5	752	2	T34355	hypothetical prote
281	31	70.5	757	2	I38423	aspartyl beta-hydr
282	31	70.5	763	2	S73178	translation initia
283	31	70.5	766	2	A53501	cyclin F - human
284	31	70.5	802	1	S48529	NAB3 protein - yea
285	31	70.5	834	2	T06055	hypothetical prote
286	31	70.5	844	2	T42702	hypothetical prote
287	31	70.5	849	1	T05181	S-receptor kinase
288	31	70.5	850	2	T14450	serine/chreonine k
289	31	70.5	857	2	S44883	ferriochrome-iron
290	31	70.5	858	2	AE2085	protein R107.4 [im
291	31	70.5	863	2	C88546	preprotein translo
292	31	70.5	867	2	H90524	hypothetical prote
293	31	70.5	869	2	T22422	hypothetical prote
294	31	70.5	872	2	S62061	SCDs protein - yea
295	31	70.5	882	2	A39030	androgen-binding p
296	31	70.5	928	1	VGEBRG	glycoprotein gi pr
297	31	70.5	928	2	S50578	hypothetical prote
298	31	70.5	929	2	T52517	hypothetical prote
299	31	70.5	932	1	VGEBRC	glycoprotein gi pr
300	31	70.5	939	2	E82121	peptidase, insulin
301	31	70.5	941	2	A86404	probable protein A
302	31	70.5	1033	2	F81595	hypothetical prote
303	31	70.5	1034	1	A33663	enteropeptidase (E
304	31	70.5	1036	1	A34755	nitrogen regulator
305	31	70.5	1043	2	C87645	AcBr/AcrD/AcrF fam
306	31	70.5	1043	2	G86550	hypothetical prote
307	31	70.5	1043	2	G72073	hypothetical prote
308	31	70.5	1046	2	T42720	cytoplasmic linker
309	31	70.5	1080	2	E87586	metal ion efflux R
310	31	70.5	1117	2	T19727	hypothetical prote
311	31	70.5	1123	2	S36846	myosin-binding pro
312	31	70.5	1132	2	S37932	hypothetical prote
313	31	70.5	1137	2	A86335	T20H2.9 protein -
314	31	70.5	1138	2	S24614	myosin-binding pro
315	31	70.5	1186	2	T23327	adenomatus polypo
316	31	70.5	1188	2	S49915	extensin-like prot
317	31	70.5	1188	2	T23330	hypothetical prote
318	31	70.5	1212	2	T00332	hypothetical prote
319	31	70.5	1265	2	T02131	hypothetical prote
320	31	70.5	1274	2	E81779	proline dehydrogen
321	31	70.5	1333	2	T00037	hypothetical prote
322	31	70.5	1325	2	T01037	hypothetical prote
323	31	70.5	1440	1	SYH1QT	multifunctional am
324	31	70.5	1445	2	T50508	hypothetical prote
325	31	70.5	1464	2	T07050	hypothetical prote
326	31	70.5	1558	2	C89114	protein C37C3.6a l
327	31	70.5	1558	2	AA47371	transcription init
328	31	70.5	2167	2	T34395	hypothetical prote
329	31	70.5	2416	2	T13825	adenomatus polypo
330	31	70.5	2763	1	M2BR22	gene 22 protein -
331	31	70.5	2774	2	A43359	microtubule-associ
332	31	70.5	26926	1	I38344	lctin, cardiac mus
333	30.5	69.3	127	2	AB2800	conserved hypothet
334	30.5	69.3	147	2	D97579	hypothetical prote
335	30.5	69.3	282	2	T26635	hypothetical prote
336	30	68.2	46	2	T46224	hypothetical prote
337	30	68.2	53	2	S31531	ribulose-bisphosph
338	30	68.2	54	2	S52452	ribulose-bisphosph
339	30	68.2	54	2	S31529	ribulose-bisphosph
340	30	68.2	54	2	S31530	ribulose-bisphosph
341	30	68.2	54	2	S31552	ribulose-bisphosph
342	30	68.2	54	2	S31527	ribulose-bisphosph
343	30	68.2	54	2	S31528	ribulose-bisphosph
344	30	68.2	54	2	S31533	ribulose-bisphosph
345	30	68.2	54	2	S31537	ribulose-bisphosph
346	30	68.2	54	2	S31536	ribulose-bisphosph
347	30	68.2	54	2	S32706	ribulose-bisphosph
348	30	68.2	55	2	S31535	ribulose-bisphosph
349	30	68.2	55	2	PQ0795	ribulose-bisphosph
350	30	68.2	55	2	PQ0797	ribulose-bisphosph
351	30	68.2	56	2	S49289	ribulose-bisphosph
352	30	68.2	56	2	S49291	ribulose-bisphosph
353	30	68.2	56	2	S49288	ribulose-bisphosph
354	30	68.2	56	2	S49274	ribulose-bisphosph
355	30	68.2	56	2	S49298	ribulose-bisphosph
356	30	68.2	56	2	S49285	ribulose-bisphosph
357	30	68.2	56	2	S49275	ribulose-bisphosph
358	30	68.2	56	2	S49277	ribulose-bisphosph
359	30	68.2	56	2	S49297	ribulose-bisphosph
360	30	68.2	56	2	S49282	ribulose-bisphosph
361	30	68.2	56	2	S49281	ribulose-bisphosph
362	30	68.2	56	2	S49278	ribulose-bisphosph
363	30	68.2	56	2	S49276	ribulose-bisphosph
364	30	68.2	56	2	S49279	ribulose-bisphosph
365	30	68.2	56	2	S49283	ribulose-bisphosph
366	30	68.2	56	2	S49284	ribulose-bisphosph
367	30	68.2	56	2	S49286	ribulose-bisphosph
368	30	68.2	56	2	S49287	ribulose-bisphosph
369	30	68.2	56	2	S49288	ribulose-bisphosph
370	30	68.2	56	2	S49290	ribulose-bisphosph
371	30	68.2	56	2	S49296	ribulose-bisphosph
372	30	68.2	57	2	S31545	ribulose-bisphosph
373	30	68.2	57	2	S31542	ribulose-bisphosph
374	30	68.2	57	2	S31540	ribulose-bisphosph
375	30	68.2	57	2	S31553	ribulose-bisphosph
376	30	68.2	57	2	S39285	ribulose-bisphosph
377	30	68.2	57	2	S39287	ribulose-bisphosph
378	30	68.2	57	2	S39288	ribulose-bisphosph
379	30	68.2	57	2	S39289	ribulose-bisphosph
380	30	68.2	57	2	S39271	ribulose-bisphosph
381	30	68.2	57	2	S39275	ribulose-bisphosph
382	30	68.2	57	2	S39276	ribulose-bisphosph
383	30	68.2	57	2	S39290	ribulose-bisphosph
384	30	68.2	57	2	S39291	ribulose-bisphosph
385	30	68.2	57	2	S52451	ribulose-bisphosph
386	30	68.2	58	2	S31547	ribulose-bisphosph
387	30	68.2	58	2	S31547	ribulose-bisphosph
388	30	68.2	58	2	S31550	ribulose-bisphosph
389	30	68.2	58	2	S31551	ribulose-bisphosph
390	30	68.2	58	2	S31541	ribulose-bisphosph
391	30	68.2	58	2	S31538	ribulose-bisphosph
392	30	68.2	58	2	S31534	ribulose-bisphosph
393	30	68.2	58	2	S31549	ribulose-bisphosph
394	30	68.2	58	2	S31548	ribulose-bisphosph
	30	68.2	64	2	S32220	ribulose-bisphosph

395	30	68.2	67	2	C81174	hypothetical prote
396	30	68.2	102	2	B34153	Ig kappa chain V-I
397	30	68.2	112	2	T51569	hypothetical prote
398	30	68.2	116	2	S33810	ribulose-bisphosph
399	30	68.2	119	2	B53482	regulatory protein
400	30	68.2	128	2	D38355	basic proline-rich
401	30	68.2	158	2	UC6056	ubiquitin-protein
402	30	68.2	166	1	PIH0SC	salivary proline-r
403	30	68.2	170	2	B25372	oligodendrocyte-sp
404	30	68.2	170	2	B55663	proline-rich phosph
405	30	68.2	171	2	A27307	probable aminoglyc
406	30	68.2	179	2	F95417	hypothetical prote
407	30	68.2	180	2	T15426	hypothetical prote
408	30	68.2	185	2	S76706	hypothetical prote
409	30	68.2	187	2	E86324	protein F14d16.27
410	30	68.2	188	2	JH0481	basic proline-rich
411	30	68.2	188	2	AF2834	conserved hypochet
412	30	68.2	195	2	B97612	hypothetical prote
413	30	68.2	207	2	T24858	hypothetical prote
414	30	68.2	211	2	S37792	hypothetical prote
415	30	68.2	221	2	G72665	hypothetical prote
416	30	68.2	227	2	E90420	DNA endonuclease I
417	30	68.2	234	1	S76823	hypothetical prote
418	30	68.2	234	2	T30427	probable apoptosis
419	30	68.2	236	2	B85025	hypothetical prote
420	30	68.2	239	2	T03078	conserved hypochet
421	30	68.2	241	2	AD2366	hypothetical prote
422	30	68.2	243	2	E70846	hypothetical prote
423	30	68.2	249	2	T10189	L-ascorbate peroxi
424	30	68.2	250	2	D86214	hypothetical prote
425	30	68.2	250	2	T08071	L-ascorbate peroxi
426	30	68.2	250	2	S43157	L-ascorbate peroxi
427	30	68.2	250	2	T07056	L-ascorbate peroxi
428	30	68.2	250	2	A45116	L-ascorbate peroxi
429	30	68.2	250	2	S49914	L-ascorbate peroxi
430	30	68.2	250	2	JE0232	L-ascorbate peroxi
431	30	68.2	251	1	PIH0PF	salivary proline-r
432	30	68.2	254	2	G72200	L-phosphoserine
433	30	68.2	263	2	S20866	L-ascorbate peroxi
434	30	68.2	268	2	T15169	hypothetical prote
435	30	68.2	269	2	T24637	hypothetical prote
436	30	68.2	273	2	C88429	protein ceh-43 [im
437	30	68.2	281	2	I38707	fas ligand - human
438	30	68.2	282	2	S43577	C28A5.4 protein (c
439	30	68.2	285	1	I46207	thioredoxin - dog
440	30	68.2	285	1	C97279	thioredoxin reduct
441	30	68.2	290	2	AC2030	hypothetical prote
442	30	68.2	293	2	D90456	oxydoreductase, pr
443	30	68.2	297	2	JQ1205	attachment protein
444	30	68.2	297	2	JQ1204	major surface glyco
445	30	68.2	298	1	MGNZ	hypothetical prote
446	30	68.2	304	2	T48281	hypothetical prote
447	30	68.2	309	2	T29293	hypothetical prote
448	30	68.2	310	1	PIH0SD	salivary proline-r
449	30	68.2	311	2	E81703	phospholipase D fa
450	30	68.2	315	2	EJ0397	DNA-directed RNA p
451	30	68.2	318	2	JC1171	muscle regulatory
452	30	68.2	322	2	S25299	extensin precursor
453	30	68.2	323	1	GHRB	ig gamma chain C r
454	30	68.2	331	2	C87385	conserved hypochet
455	30	68.2	336	1	BYECPR	phosphatase-repress
456	30	68.2	346	1	AB0956	periplasmic phosph
457	30	68.2	347	2	S36980	hypothetical prote
458	30	68.2	349	1	QOBE05	HHRF5 protein - hu
459	30	68.2	350	2	I38403	neu differentiation
460	30	68.2	350	2	T04097	acyl-[acyl]-carrier
461	30	68.2	352	1	PIH0B6	salivary proline-r
462	30	68.2	356	2	T14172	acyl-[acyl]-carrier
463	30	68.2	356	2	T14264	acyl-[acyl]-carrier
464	30	68.2	356	2	T14268	acyl-[acyl]-carrier
465	30	68.2	396	2	S24995	acyl-[acyl]-carrier
466	30	68.2	401	2	E84869	stearyl-ACP desat

468	30	68.2	401	2	T02958	ribulose-bisphosph
469	30	68.2	402	2	A40678	T-cell adhesion re
470	30	68.2	404	2	S68409	potassium channel
471	30	68.2	406	2	T44819	dolichyl-phosphate
472	30	68.2	407	2	S71264	acyl-[acyl]-carrier
473	30	68.2	411	2	T07806	acyl-[acyl]-carrier
474	30	68.2	411	2	C36793	hypothetical prote
475	30	68.2	413	2	AH2743	conserved hypochet
476	30	68.2	416	2	G97524	hypothetical prote
477	30	68.2	424	2	E83442	probable MFS trans
478	30	68.2	426	1	RKHLC	ribulose-bisphosph
479	30	68.2	427	1	GQHN	nerve growth facto
480	30	68.2	427	1	A13490	membrane-bound lyl
481	30	68.2	431	2	T01557	hypothetical prote
482	30	68.2	443	1	T08339	transcription fact
483	30	68.2	448	2	F81703	phospholipase D fa
484	30	68.2	450	1	G72039	2-amino-4-hydroxy-
485	30	68.2	450	2	D86585	dihydropterolate by
486	30	68.2	452	2	S47228	ribulose-bisphosph
487	30	68.2	452	2	S47229	ribulose-bisphosph
488	30	68.2	452	2	S47230	ribulose-bisphosph
489	30	68.2	452	2	S47231	ribulose-bisphosph
490	30	68.2	452	2	S47232	ribulose-bisphosph
491	30	68.2	452	2	S47233	ribulose-bisphosph
492	30	68.2	452	2	S47234	ribulose-bisphosph
493	30	68.2	452	2	S47235	ribulose-bisphosph
494	30	68.2	452	2	S47236	ribulose-bisphosph
495	30	68.2	452	2	S47237	ribulose-bisphosph
496	30	68.2	452	2	S47238	ribulose-bisphosph
497	30	68.2	452	2	S47221	ribulose-bisphosph
498	30	68.2	452	2	S47222	ribulose-bisphosph
499	30	68.2	452	2	S47223	ribulose-bisphosph
500	30	68.2	452	2	S47224	ribulose-bisphosph

## ALIGNMENTS

## RESULT 1

UKHU

u-plasminogen activator (BC 3.4.21.73) precursor [validated] - human

N;Alternate names: cellular plasminogen activator; urokinase; urokinase-type plasminogen

N;Contents: urokinase-type plasminogen activator chain A; urokinase-type plasminogen acti

in form

C;Species: Homo sapiens (man)

C;Date: 17-Dec-1982 #sequence revision 04-Dec-1986 #text change 15-Sep-2000

C;Accession: A00931; 152209; J0102; A37561; 138102; S65783; A37563; A37564; A356

R;Riccio, A.; Grimaldi, G.; Verde, P.; Sebastio, G.; Boast, S.; Blasi, F.

Nucleic Acids Res. 13, 2759-2771, 1985

A;Title: The human urokinase-plasminogen activator gene and its promoter.

A;Reference number: A00931; WUID:85215647; PMID:2387867

A;Accession: A00931

A;Molecule type: DNA

A;Residues: 1431 &lt;RIC&gt;

A;Cross-references: GB:K02419; NID:G37601; PIDN:CAA26268.1; PID:G1834524

A;Note: the authors translated the codon ATG for residue 214 as Ile

R;Maganine, Y.; Pearson, D.; Gratian, M.

Biochem. Biophys. Res. Commun. 132, 563-569, 1985

A;Title: Exon-intron boundary sliding in the generation of two mRNAs coding for porcine l

A;Reference number: 152209; WUID:86050639; PMID:3933505

A;Accession: 152209

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 145-161 &lt;NMG1&gt;

A;Cross-references: GB:K03027; NID:G340174; PIDN:AAA61257.1; PID:G340175

R;Negai, M.; Hiramatsu, K.; Kaneda, T.; Hayasuke, N.; Arimura, H.; Nishida, M.; Suyama, J

Gene 36, 183-189, 1985

A;Title: Molecular cloning of cDNA coding for human preprourokinase.

A;Reference number: J0102; WUID:86056954; PMID:2415429

A;Accession: J0102

A;Molecule type: mRNA

A;Residues: 1-213; T, 215-431 &lt;NMG2&gt;

A;Cross-references: GB:K03226; NID:G340155; PIDN:AAC97138.1; PID:G340158; GB:D00244; NID

R.Verde, P.; Stoppelli, M.P.; Galeffi, P.; Di Nocera, P.; Biasi, F.  
 Proc. Natl. Acad. Sci. U.S.A. 81, 4727-4731, 1984  
 A>Title: Identification and primary sequence of an unspliced human urokinase poly(A) + RN  
 A:Reference number: A37561; PMID:84272706; PMID:6589620  
 A:Accession: A37561  
 A:Molecule type: mRNA  
 A:Residues: 66-431 <VER>  
 A:Cross-references: GB:D00244; NID:9230138  
 R.Jacobs, P.; Cravador, A.; Lorian, R.; Brockly, F.; Colau, B.; Chuchana, P.; van Elsen,  
 DNA 4, 139-146, 1985  
 A>Title: Molecular cloning, sequencing, and expression in *Escherichia coli* of human pre  
 A:Reference number: I38102; PMID:85203359; PMID:3888571  
 A:Accession: I38102  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-150, 'W', 152-213, 'T', 215-385, 'C', 387-429, 'V', 431 <JAC>  
 A:Cross-references: EMBL:X02760; NID:935297; PID:CA26535.1; PID:935298  
 R.Yoshimoto, M.; Ushiyama, Y.; Sakai, M.; Tamaki, S.; Hara, H.; Takahashi, K.; Sawasaki,  
 Biochim. Biophys. Acta 1293, 83-89, 1996  
 A>Title: Characterization of single chain urokinase-type plasminogen activator with a nd  
 A:Reference number: S65783; PMID:86186279; PMID:8652631  
 A:Accession: S65783  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 21-140, 'L', 142-213, 'T', 215-431 <YOS>  
 A:Cross-references: EMBL:D1143; NID:9311467; PID:BA01919.1; PID:93119928  
 R.Gunzler, W.A.; Steffens, G.J.; Octing, F.; Kim, S.M.A.; Frankus, E.; Flohe, L.  
 Hoppe-Seyler's Z. Physiol. Chem. 363, 1155-1165, 1982  
 A>Title: The primary structure of high molecular mass urokinase from human urine.  
 A:Reference number: A37562; PMID:83055084; PMID:6754569  
 A:Accession: A37562  
 A:Molecule type: protein  
 A:Residues: 21-177 <GUN>  
 R.Schaller, J.; Nick, H.; Rickli, E.E.; Gillesen, D.; Lergier, W.; Studer, R.O.  
 Eur. J. Biochem. 125, 251-257, 1982  
 A>Title: Human low-molecular-weight urinary urokinase. Partial characterization and prel  
 A:Reference number: A37563; PMID:83003608; PMID:6749491  
 A:Accession: A37563  
 A:Molecule type: protein  
 A:Residues: 156-176, 179-193, 'T', 195, 'T', 197-224 <SCH>  
 R.Steffens, G.J.; Gunzler, W.A.; Octing, F.; Frankus, E.; Flohe, L.  
 Hoppe-Seyler's Z. Physiol. Chem. 363, 1043-1058, 1982  
 A>Title: The complete amino acid sequence of low molecular mass urokinase from human uri  
 A:Reference number: A37564; PMID:83055099; PMID:6754572  
 A:Accession: A37564  
 A:Molecule type: protein  
 A:Residues: 158-410 <STE>  
 R.Kentzer, E.J.; Buko, A.; Menon, G.; Sarin, V.K.  
 Biochem. Biophys. Res. Commun. 171, 401-406, 1990  
 A>Title: Carbohydrate composition and presence of a fucose-protein linkage in recombinat  
 A:Reference number: A35689; PMID:90365737; PMID:2393398  
 A:Accession: A35689  
 A:Molecule type: protein  
 A:Residues: 21-30, 'X', 32, 'X', 34-38, 'X', 40-43 <KEN>  
 A:Notes: identification of a fucose and attempt to determine its attachment site  
 R.Rabani, S.A.; Desjardins, J.; Bell, A.W.; Benville, D.; Mazar, A.; Henkin, J.; Goltz  
 Biochem. Biophys. Res. Commun. 173, 1056-1064, 1990  
 A>Title: An amino-terminal fragment of urokinase isolated from a prostate cancer cell li  
 A:Reference number: A36697; PMID:91097529; PMID:2125213  
 A:Accession: A36697  
 A:Molecule type: protein  
 A:Residues: 21-34 <RAB>  
 R.Li, X.; Bokman, A.M.; Linae, M.; Smith, R.A.G.; Dobson, C.M.  
 submitted to the Brookhaven Protein Data Bank, July 1993  
 A:Reference number: A51255; PDB:1XDU  
 A:Contents: annotation; conformation and disulfide bond assignments by (1)H-NMR, residu  
 R.Li, X.; Smith, R.A.G.; Dobson, C.M.  
 Biochemistry 31, 9562-9571, 1992  
 A>Title: Sequential (1)H NMR assignments and secondary structure of the kringe domain f  
 A:Reference number: A43475; PMID:33003110; PMID:1327118  
 A:Contents: annotation; conformation and disulfide bond assignments by (1)H-NMR  
 R.Hansen, A.P.; Petros, A.M.; Meadows, R.P.; Nettesheim, D.G.; Mazar, A.P.; Olejniczak,  
 submitted to the Brookhaven Protein Data Bank, January 1994

A:Reference number: A66822; PDB:1URK  
 A:Contents: annotation; conformation and disulfide bond assignments by (1)H-NMR, residues  
 R.Sprengon, G.S.; Phillips, C.; Nowak, U.K.; Ponting, C.P.; Saunders, D.; Dobson, C.M.;  
 submitted to the Brookhaven Protein Data Bank, July 1995  
 A:Reference number: A66058; PDB:1LMW  
 A:Contents: annotation; X-ray crystallography, 2.5 angstroms, residues 168-175, 179-426  
 C:Comment: This enzyme is found in urine in a high molecular mass form, consisting of A  
 C:Comment: Urokinase-type plasminogen activator proteolytically activates plasminogen, ar  
 C:Genetics: GDB:PLAU  
 A:Gene: GDB:PLAU  
 A:Cross-references: GDB:119497; OMIM:191840  
 A:Map position: 10q24-10q24  
 A:Introns: 19/3; 29/1; 65/1; 123/2; 154/1; 227/2; 277/1; 324/1; 373/3  
 C:Function:  
 A>Description: proteolytically activates plasminogen  
 A:Pathway: Fibrinolysis  
 C:Superfamily: urokinase-type plasminogen activator; EGF homology; kringe homology; try  
 C:Keywords: fibrinolysis; glycoprotein; heterodimer; hydrolase; kringe; serine proteina  
 F:1-20/Domain: signal sequence #status predicted <Sig>  
 F:21-431/Product: urokinase-type plasminogen activator, single chain form #status predict  
 F:21-177/Product: urokinase-type plasminogen activator, chain A #status experimental <MP  
 F:31-62/Domain: EGF homology <EGF>  
 F:70-151/Domain: kringe homology <KRG>  
 F:156-177/Product: urokinase-type plasminogen activator chain A1 #status experimental <M  
 F:179-431/Product: urokinase-type plasminogen activator chain B #status experimental <MP  
 F:179-419/Domain: trypsin homology <TRY>  
 F:31-39,33-51,53-62,70-151,91-133,122-166,168-299,209-225,217-288,313-382,345-361,372-40  
 F:38/Binding site: carbohydrate (Thr) (covalent) #status predicted  
 F:178-179/Cleavage site: Lys-116 (plasmin) #status experimental  
 F:224,275,376/Active site: His, Asp, Ser #status experimental  
 F:322/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 100.0%; Score 44; DB 1; Length 433;  
 Best Local Similarity 100.0%; Pred. No. 7.6;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 156 KPSPPPEE 163  
 |||||  
 |||||

## RESULT 2

U-plasminogen activator (EC 3.4.21.73) precursor - yellow baboon  
 C:Species: Papio cynocephalus, Papio hamadryas cynocephalus (yellow baboon)  
 C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 18-Jun-1999  
 C:Accession: S14687; S08651  
 R:Aw, Y.P.T.; Wang, T.W.; Clowes, A.W.  
 Nucleic Acids Res. 18, 3411, 1990  
 A>Title: Nucleotide and deduced amino acid sequences of baboon urokinase-type plasminoge  
 A:Reference number: S14687; PMID:90287734; PMID:2113276  
 A:Accession: S14687  
 A:Molecule type: mRNA  
 A:Residues: 1-433 <AW>  
 A:Cross-references: EMBL:X51935; NID:938130; PID:CA36200.1; PID:938131  
 C:Superfamily: urokinase-type plasminogen activator; EGF homology; kringe homology; try  
 C:Keywords: glycoprotein; heterodimer; hydrolase; kringe; serine proteinase  
 F:1-20/Domain: signal sequence #status predicted <Sig>  
 F:21-176/Product: plasminogen activator chain A #status predicted <ACH>  
 F:30-61/Domain: EGF homology <EGF>  
 F:69-150/Domain: kringe homology <KRG>  
 F:178-433/Product: plasminogen activator chain B #status predicted <BCH>  
 F:178-421/Domain: trypsin homology <TRI>  
 F:167-298,208-224,216-287,315-384,347-363,374-402/Disulfide bonds: #status predicted  
 F:223,274,378/Active site: His, Asp, Ser #status predicted  
 F:324/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 44; DB 1; Length 433;  
 Best Local Similarity 100.0%; Pred. No. 7.6;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KPSPPPEE 8  
 |||||



R:Lamerdin, J.E.; McCready, P.M.; Skowronski, E.; Adamson, A.W.; Burkhardt-Schultz, K.; Christensen, M.; Georgescu, A.; Avila, J.; Liu, S.; Attix, C.; Andreise, T.; Trankiel submitted to the EMBL Data Library, March 1998  
 A:Authors: Duarte, S.; Lucas, S.; Bruce, R.; Thomas, P.; Quan, G.; Krommiller, B.; Arell A:Description: Sequence analysis of a human PI clone containing the XCC9 DNA repair gen A:Reference number: Z14637  
 A:Accession: T02245  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-880 <LM>  
 A:Cross-references: EMBL:AC004472; NID:G2984582; PID:AC07985.1; PID:G2984587 C:Genetics:  
 A:Map position: 9  
 A:Introns: 89/3; 152/1; 200/1; 241/2; 294/3; 354/3; 435/1; 733/1; 802/1; 871/1

Query Match 81.8%; Score 36; DB 2; Length 880;  
 Best Local Similarity 85.7%; Pred. No. 3.1e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PSSPPE 8  
 Db 289 PSSPPE 295

RESULT 8  
 T46428  
 hypothetical protein DKFZp434B2226.1 - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
 C:Accession: T46428  
 R:Angeorge, W.; Wixner, U.; Mewes, H.W.; Gaassenhuber, J.; Wiemann, S. submitted to the Protein Sequence Database, January 2000  
 A:Reference number: Z23028  
 A:Accession: T46428  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1185 <AA>  
 A:Cross-references: EMBL:AL137357  
 A:Experimental source: adult testis; clone DKFZp434B2226  
 C:Genetics:  
 A:Note: DKFZp434B2226.1

Query Match 81.8%; Score 36; DB 2; Length 1185;  
 Best Local Similarity 85.7%; Pred. No. 4.2e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PSSPPE 8  
 Db 578 PSSPPE 584

RESULT 9  
 A29360  
 SIR4 protein - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: protein YD934.12; protein YDR227W; STE9 protein  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 21-Jul-2000  
 C:Accession: A29360; S47935; S59434; S53988; S47476; S47950  
 R:Marshall, M.; Mahoney, D.; Rose, A.; Hicke, J.B.; Broach, J.R. Mol. Cell. Biol. 7, 4441-4452, 1987  
 A:Title: Functional domains of SIR4, a gene required for position effect regulation in S A:Reference number: A29360; NID:88142836; PMID:3325825  
 A:Accession: A29360  
 A:Molecule type: DNA  
 A:Residues: 1-1358 <MAR>  
 A:Cross-references: GB:M37249; NID:G531115; PID:AAA20881.1; PID:G531116  
 R:Davies, C.J.; Hutchison III, C.A. submitted to the EMBL Data Library, September 1994  
 A:Description: Tm3 transposon/deletion sequencing of a 9.4kb DNA fragment. Characterist A:Accession: S47935  
 A:Reference number: S47932  
 A:Molecule type: DNA  
 A:Residues: 1-993; 'L', 995-1358 <DAV>

A:Cross-references: EMBL:LJ5344; EMBL:U13239; EMBL:Z36548  
 R:Murphy, L.; Harris, D. submitted to the EMBL Data Library, March 1995  
 A:Reference number: S59423  
 A:Accession: S59434  
 A:Molecule type: DNA  
 A:Residues: 1-1358 <MUR>  
 A:Cross-references: EMBL:Z48612; NID:G728671; PID:G728683; MIPS:YDR227W  
 A:Experimental source: strain AB972  
 R:Davies, C.J.; Hutchison III, C.A. Nucleic Acids Res. 23, 507-514, 1995  
 A:Title: Insertion site specificity of the transposon Tm3.  
 A:Reference number: S53985; NID:95192063; PMID:7685847  
 A:Accession: S53988  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-993; 'L', 995-1358 <DAV>  
 A:Cross-references: EMBL:U13239; NID:G532747; PID:AA03144.1; PID:G532751  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994  
 C:Genetics:  
 A:Gene: SGD:STR4; STE9  
 A:Cross-references: SGD:S0002635; MIPS:YDR227W  
 A:Map position: 4R  
 C:Keywords: DNA binding; transcription regulation

Query Match 81.8%; Score 36; DB 2; Length 1358;  
 Best Local Similarity 85.7%; Pred. No. 4.8e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPE 7  
 Db 1131 KPSSPPE 1137

RESULT 10  
 CGHUIE  
 collagen alpha 1(XI) chain precursor - human  
 N:Alternate names: procollagen alpha 1(XI) chain  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Mar-1990 #sequence\_revision 03-Oct-1995 #text\_change 08-May-1998  
 C:Accession: A35239; A31795  
 R:Yoshioka, H.; Ramirez, F. J. Biol. Chem. 265, 6423-6426, 1990  
 A:Title: Pro-alpha1(XI) collagen. Structure of the amino-terminal propeptide and expressi A:Reference number: A35239; NID:90202924; PMID:1690726  
 A:Accession: A35239  
 A:Molecule type: mRNA  
 A:Residues: 1-558 <YOS>  
 A:Cross-references: GB:J05407  
 R:Bernard, M.; Yoshioka, H.; Rodriguez, E.; van der Rest, M.; Kimura, T.; Ninomiya, Y.; C U. Biol. Chem. 263, 17159-17166, 1988  
 A:Title: Cloning and sequencing of pro-alpha1(XI) collagen cDNA demonstrates that type XI cartilaginous tissue.  
 A:Reference number: A92689; NID:89034222; PMID:3182841  
 A:Accession: A31795  
 A:Molecule type: DNA; mRNA  
 A:Residues: 538-1806 <BER>  
 A:Cross-references: GB:J04177  
 A:Note: parts of this sequence were determined by protein sequencing  
 C:Comment: Prolins and lysines at the third position of the tripeptide repeating unit (C ed and subsequently O-glycosylated).  
 C:Genetics:  
 A:Gene: GDB:COL11A1; COL16  
 A:Cross-references: GDB:I20595; OMIM:120280  
 A:Map position: 1p21-1p21  
 A:Introns: 561/3; 579/3; 615/3; 633/3; 648/3; 666/3; 681/3  
 A:Note: the list of introns is incomplete  
 C:Complex: type XI collagen may be a heterotrimer of two alpha 1(XI) chains and one alpha 3(XI) chain (see PIR:G0H6C), initially linked by disulfide bonds among their carboxyl-termed with desmosine cross-links made from lysine and allysine residues  
 C:Function:  
 A:Description: structural component of extracellular fibrous polymer associated with cell A:Note: may play a role in controlling the lateral growth of collagen II fibrils

C:Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology  
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline  
 F:1-36/Domain: signal sequence #status predicted <SIG>  
 F:35-260/Domain: PAMP-like #status predicted <PAMP>  
 F:37-511/Domain: amino-terminal propeptide #status predicted <PRO>  
 F:512-1565/Product: collagen alpha 1(XI) chain #status predicted <MAT>  
 F:512-527/Region: amino-terminal nonhelical telopeptide  
 F:528-1542/Region: helical  
 F:1543-1565/Region: carboxyl-terminal nonhelical telopeptide  
 F:1566-1806/Domain: carboxyl-terminal propeptide #status predicted <CTP>  
 F:1583-1805/Domain: fibrillar collagen carboxyl-terminal homology <FCC>  
 F:51-243,182-236/Disulfide bonds: #status predicted  
 F:505/Modified site: allylsine (Lys) #status predicted  
 F:512,1452/Modified site: 5-hydroxylysine (Lys) #status predicted  
 F:512,1452/Binding site: carbohydrate (Lys) (covalent) #status predicted

Query Match 81.8%; Score 36; DB 1; Length 1806;  
 Best Local Similarity 75.0%; Pred. No. 6.4e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KPSPPEE 8  
 ||:||||  
 Db 392 KPSPPEE 399

RESULT 11  
 B83542  
 hypothetical protein PA0818 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa  
 C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: B83542  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mitsuoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
 .; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 gen  
 A:Reference number: A82950; MUID:20437337; PMID:10984043  
 A:Accession: B83542  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-90 <STO>  
 A:Cross-references: GB:AE004517; GB:AE004091; NID:9946710; PIDN:AAC04207.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA0818

Query Match 79.5%; Score 35; DB 2; Length 90;  
 Best Local Similarity 75.0%; Pred. No. 45;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KPSPPEE 8  
 ||:||||  
 Db 44 KPSPPEE 51

RESULT 12  
 T26841

hypothetical protein Y43F4B.3 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T26841  
 R:Matthews, L.  
 submitted to the EMBL Data Library, January 1998  
 A:Reference number: Z20276  
 A:Accession: T26841  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-216 <WTL>  
 A:Cross-references: EMBL:AL021481; PIDN:CAA16332.1; CESP:Y43F4B.3  
 A:Experimental source: clone Y43F4B  
 C:Genetics:  
 A:Gene: CESP:Y43F4B.3  
 A:Introns: 59/2

Query Match 79.5%; Score 35; DB 2; Length 216;  
 Best Local Similarity 75.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KPSPPEE 8  
 ||:||||  
 Db 68 KPSPPEE 75

RESULT 13  
 H72858

apoptosis inhibitor - Autographa californica nuclear polyhedrosis virus  
 C:Species: Autographa californica nuclear polyhedrosis virus, AcMNPV  
 A:Note: dsDNA virus  
 C>Date: 12-Nov-1999 #sequence\_revision 12-Nov-1999 #text\_change 12-Nov-1999  
 C:Accession: H72858  
 R:Ayres, M.D.; Howard, S.C.; Kuzio, J.; Lopez-Ferber, M.; Possee, R.D.  
 Virology 202, 586-605, 1994  
 A>Title: The complete DNA sequence of Autographa californica nuclear polyhedrosis virus.  
 A:Reference number: A72850; MUID:94303173; PMID:8030224  
 A:Accession: H72858  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-249 <AIR>  
 A:Cross-references: GB:I22858; NID:9510708; PIDN:AAA66701.1; PID:9559140  
 C:Genetics:  
 A:Gene: Ac-IAP2

Query Match 79.5%; Score 35; DB 2; Length 249;  
 Best Local Similarity 75.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KPSPPEE 8  
 ||:||||  
 Db 186 KPSPPEE 193

RESULT 14  
 T41814

IAP2 orf71 - Bombyx mori nuclear polyhedrosis virus (isolate T3)  
 C:Species: Bombyx mori nuclear polyhedrosis virus, BMSNPV  
 A:Variety: isolate T3  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 11-May-2000  
 C:Accession: T41814  
 R:Gomi, S.; Majima, K.; Maeda, S.  
 J. Gen. Virol. 80, 1323-1337, 1999  
 A>Title: Sequence analysis of the genome of Bombyx mori nucleopolyhedrovirus.  
 A:Reference number: Z22020; MUID:99281911; PMID:10355780  
 A:Accession: T41814  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-249 <RAM>  
 A:Cross-references: EMBL:U3180; PIDN:AAC63743.1  
 A:Experimental source: isolate T3  
 C:Genetics:  
 A:Note: IAP2

Query Match 79.5%; Score 35; DB 2; Length 249;  
 Best Local Similarity 75.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KPSPPEE 8  
 ||:||||  
 Db 186 KPSPPEE 193

RESULT 15  
 T07887

G box-binding protein homolog GBF1 - rape  
 C:Species: Brassica napus (rape)  
 C>Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 26-May-2000  
 C:Accession: T07887



R;Bellemare, G.  
submitted to the EMBL Data Library, May 1995  
A:Reference number: Z16192  
A:Accession: T07887  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-374 <BEL>  
A:Cross-references: EMBL:U27108; NID:g1399006; PIDN:AA03379.1; PID:g1399007  
A:Experimental source: cv. Westar; roots from mature plants  
C:Genetics:  
A:Gene: GBF1  
A:Introns: 21/3  
C:Superfamily: fava bean G-box-binding protein; fos/jun DNA-binding domain homology  
C:Keywords: DNA binding; leucine zipper; transcription factor; transcription regulation  
F:253-293/Domain: fos/jun DNA-binding domain homology <F0D>

Query Match 79.5%; Score 35; DB 2; Length 374;  
Best Local Similarity 75.0%; Pred. No. 1.9e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

RESULT 16  
T07882  
G box-binding protein homolog GBF2 - rape  
C:Species: Brassica napus (rape)  
C>Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 26-May-2000  
A:Accession: T07882  
R;Bellemare, G.  
submitted to the EMBL Data Library, May 1995  
A:Reference number: Z16192  
A:Accession: T07882  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-376 <BEL>  
A:Cross-references: EMBL:U27107; NID:g1399004; PIDN:AA03378.1; PID:g1399005  
A:Experimental source: cv. Westar; roots from mature plants  
C:Genetics:  
A:Introns: 21/3; 36/3; 69/3; 97/3; 109/3; 175/3; 192/1; 256/3; 282/3; 324/3  
A>Note: BnGBF2  
C:Superfamily: fava bean G-box-binding protein; fos/jun DNA-binding domain homology  
C:Keywords: DNA binding; leucine zipper; transcription factor; transcription regulation  
F:253-295/Domain: fos/jun DNA-binding domain homology <F0D>

Query Match 79.5%; Score 35; DB 2; Length 376;  
Best Local Similarity 75.0%; Pred. No. 1.9e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KPSSPBE 8  
DB 15 KPSSPAD 22

RESULT 17  
D69316  
mRNA 3'-end processing factor homolog - Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 21-Jul-2000  
A:Accession: D69316  
R;Flenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A:Authors: Uetacker, T.; Cotton, M.D.; Spriggs, T.; Artlach, P.; Kaine, B.P.; Sykes, S.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon  
A:Reference number: A69250; MUID:98049343; PMID:9389475  
A:Accession: D69316  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA

A:Residues: 1-407 <KLE>  
A:Cross-references: GB:AE001067; GB:AE000782; NID:g2689390; PIDN:AA09702.1; PID:g265008

Query Match 79.5%; Score 35; DB 2; Length 407;  
Best Local Similarity 85.7%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KPSSPBE 7  
DB 30 KPSSDPE 36

RESULT 18  
I50213  
protein-tyrosine-phosphatase (EC 3.1.3.48) - chicken (fragment)  
C:Species: Gallus gallus (chicken)  
C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jan-2000  
A:Accession: I50213  
R;Stoker, A.W.  
Mach. Dev. 46, 201-217, 1994  
A:Title: Isoforms of a novel cell adhesion molecule-like protein tyrosine phosphatase are  
A:Reference number: I50212; MUID:95001563; PMID:7918104  
A:Accession: I50213  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-440 <STO>  
A:Cross-references: GB:I32781; NID:g485748; PIDN:AA64461.1; PID:g485749  
C:Genetics:  
A:Gene: CYPalpha2  
C:Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology; 3  
ogy  
C:Keywords: phosphoric monoester hydrolase; tyrosine-specific phosphatase  
F:123-214/Domain: fibronectin type III repeat homology <F0R>

Query Match 79.5%; Score 35; DB 2; Length 440;  
Best Local Similarity 62.5%; Pred. No. 2.2e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPBE 8  
DB 20 KPSPPOD 27

RESULT 19  
E96583  
hypothetical protein F20D21.4 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Nov-2001  
A:Accession: E96583  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Com, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hultzer, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Matli, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I. ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: E96583  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-516 <STO>  
A:Cross-references: GB:AE005173; NID:g4585966; PIDN:AA025602.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F20D21.4  
A:Map position: 1  
C:Superfamily: dihydroliposamide acetyltransferase; lipoyl/biotin-binding homology

Query Match 79.5%; Score 35; DB 2; Length 516;  
Best Local Similarity 85.7%; Pred. No. 2.6e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSSPPE 7  
:|||||  
Db 203 QPSSPPE 209

## RESULT 20

BABOH

peptide-aspartate beta-dioxygenase (EC 1.14.11.16) - bovine  
N:Alternate names: aspartyl (asparaginyl) beta-hydroxylase  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 31-Dec-1993 #sequence\_revision 10-Feb-1995 #text\_change 11-Jun-1999  
C/Accession: A42969; A39470; B39470; C39470; S27948  
R:Jia, S.; Vandusen, W.J.; Diehl, R.E.; Kohl, N.E.; Dixon, R.A.; Elliston, K.O.; Stern, J. Biol. Chem. 267, 14322-14327, 1992  
A>Title: cDNA cloning and expression of bovine aspartyl (asparaginyl) beta-hydroxylase.  
A:Reference number: A42969; PMID:2232546; PMID:1378441  
A/Accession: A42969  
A/Molecule type: mRNA  
A/Residues: 1-754 <UIA>  
A/Cross-references: EMBL:M91213; NID:G162693; PIDN:AAA03563.1; PID:G162694  
A:Experimental source: brain  
A>Note: sequence extracted from NCBI backbone (NCBI:P108534)  
R:Kang, Q.; Vandusen, W.J.; Petrocki, C.J.; Garsky, V.M.; Stern, A.M.; Friedman, P.A. J. Biol. Chem. 266, 14004-14010, 1991  
A>Title: Bovine liver aspartyl beta-hydroxylase. Purification and characterization.  
A:Reference number: A39470; PMID:91310689; PMID:1856229  
A/Accession: A39470  
A/Molecule type: protein  
A/Residues: 289-328 <MAN>  
A/Accession: B39470  
A/Molecule type: protein  
A/Residues: 615,'X', 617-630,'XX', 633-634,'X', 636,'XX', 639-641 <MA2>  
A/Accession: C39470  
A/Molecule type: protein  
A/Residues: 311-347,'X', 349,'X', 351-373,'X', 375-379,'X', 381-382 <MA3>  
C/Comment: This enzyme uses ferrous iron as a cofactor, and while beta-hydroxylating the C/Comment: Aspartic acid and asparagine residues in the BGF homology domain of certain P C/Superfamily: peptide-aspartate beta-dioxygenase; tetratricopeptide repeat homology  
C/Keywords: glycoprotein; oxidoreductase; transmembrane protein  
F:2-56/Domain: intracellular #status predicted <IN>  
F:57-78/Domain: transmembrane #status predicted <TM>  
F:289-754/Product: peptide-aspartate beta-dioxygenase, 56k form #status predicted <56k>  
F:311-754/Product: peptide-aspartate beta-dioxygenase, 52k form #status predicted <52k>  
F:337-370/Domain: tetratricopeptide repeat homology <TT1>  
F:371-404/Domain: tetratricopeptide repeat homology <TT2>  
F:13, 96, 466, 702/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 79.5%; Score 35; DB 1; Length 754;  
Best Local Similarity 75.0%; Pred. No. 3.8e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KPSSPPE 8  
:|||||  
Db 134 KPVPPE 141

RESULT 21  
A53047  
6-phosphofructokinase (EC 2.7.1.11) - rat  
C/Species: Rattus norvegicus (Norway rat)  
C/Date: 02-Jun-1995 #sequence\_revision 02-Jun-1995 #text\_change 12-May-2003  
C/Accession: A53047  
R:Cekakis, N.; Johnson, R.C.; Jenkins, A.; Maine, R.E.; Sul, H.S. J. Biol. Chem. 269, 3348-3355, 1994  
A>Title: Structure, distribution, and functional expression of the phosphofructokinase C/Accession: A53047; MUID:94148828; PMID:8106374  
A/Reference number: A53047  
A/Molecule type: preliminary  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-765 <GK>  
A/Cross-references: GB:I25387; NID:G466512; PIDN:AAA17757.1; PID:G473914  
A/Note: authors translated the codon CCG for residue 220 as Asp

C/Superfamily: 6-phosphofructokinase, eukaryotic type: 6-phosphofructokinase 1 homology  
C/Keywords: ATP; phosphofructokinase 1 homology <6PF1>  
F:4-313/Domain: 6-phosphofructokinase 1 homology <6PF1>  
F:390-676/Domain: 6-phosphofructokinase 1 homology <6PF2>

Query Match 79.5%; Score 35; DB 2; Length 765;  
Best Local Similarity 85.7%; Pred. No. 3.9e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 PSSPPE 8  
:|||||  
Db 217 PSSPPE 223

## RESULT 22

JC2055

6-phosphofructokinase (EC 2.7.1.11), platelet - human  
N:Alternate names: ATP-D-fructose-6-phosphate 1-phosphotransferase  
C/Species: Homo sapiens (man)  
C/Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 12-May-2003  
C/Accession: JC2055; JH0477  
R:Ero, K.; Sakura, H.; Yasuda, K.; Hayakawa, T.; Kawasaki, E.; Moriuchi, R.; Nagataki, S Biochem. Biophys. Res. Commun. 198, 990-998, 1994  
A>Title: Cloning of a complete protein-coding sequence of human platelet-type phosphofructokinase.  
A:Reference number: JC2055; MUID:94161770; PMID:8117307  
A/Accession: JC2055  
A/Molecule type: mRNA  
A/Residues: 1-784 <ETO>  
A/Cross-references: DDBJ:D25328; NID:G464186; PIDN:BA04998.1; PID:G560105  
A:Experimental source: pancreatic islet  
R:Simpson, C.J.; Fothergill-Gilmore, L.A. Biochem. Biophys. Res. Commun. 180, 197-203, 1991  
A>Title: Isolation and sequence of a cDNA encoding human platelet phosphofructokinase.  
A:Reference number: JH0477; MUID:92028938; PMID:1834056  
A/Accession: JH0477  
A/Molecule type: mRNA  
A/Residues: 1'P', 486-497, 499-698, 'E', 700-784 <SIM>  
A/Cross-references: GB:I64784; NID:G189852; PIDN:AAA6435.1; PID:G189853  
A:Experimental source: lymphocyte, Raji cell line  
C/Comment: This enzyme catalyzes the phosphorylation of fructose 6-phosphate to fructose C/Genes:  
A/Genes: GDB:PFKP  
A/Cross-references: GDB:119480; OMIM:171840  
A/Map position: 10p15.3-10p15.2  
C/Superfamily: 6-phosphofructokinase, eukaryotic type: 6-phosphofructokinase 1 homology  
C/Keywords: allosteric regulation; ATP; duplication; glycolysis; magnesium; phosphotransf F:27-336/Domain: 6-phosphofructokinase 1 homology <6PF1>  
F:413-699/Domain: 6-phosphofructokinase 1 homology <6PF2>  
F:203,234,236,252,430,434,468,688/Binding site: ADP/AMP, allosteric (Thr, Ala, Trp, F:567,625/Binding site: ATP (lys) #status predicted

Query Match 79.5%; Score 35; DB 2; Length 784;  
Best Local Similarity 85.7%; Pred. No. 4e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 PSSPPE 8  
:|||||  
Db 240 PSSPPE 246

RESULT 23  
A53206  
6-phosphofructokinase (EC 2.7.1.11) C - rabbit  
N:Alternate names: phosphofructo-1-kinase C  
C/Species: Oryctolagus cuniculus (domestic rabbit)  
C/Date: 19-Oct-1995 #sequence\_revision 19-Oct-1995 #text\_change 12-May-2003  
C/Accession: A53206; S03880  
R:Li, Y.; Valaitis, A.P.; Latschew, S.P.; Kwiatkowska, D.; Tripathi, R.L.; Campbell, M.C., J. Biol. Chem. 269, 5781-5787, 1994  
A>Title: Structure and expression of the cDNA for the C isozyme of phosphofructo-1-kinase A/Reference number: A53206; MUID:94164929; PMID:8119919  
A/Accession: A53206  
A/Status: preliminary

A:Molecule type: mRNA  
 A:Residues: 1-791 <L1A>  
 A:Cross-references: GB:U01154  
 R:Valaitis, A.P.; Foe, L.G.; Kwiatkowska, D.; Latehaw, S.P.; Kemp, R.G.  
 Biochim. Biophys. Acta 995, 187-194, 1999  
 A:Title: The sites of phosphorylation of rabbit brain phosphofructo-1-kinase by cyclic A  
 A:Reference number: S03878; PMID:89194250; PMID:2539199  
 A:Accession: S03880  
 A:Molecule type: protein  
 A:Residues: 5-17 <VAL>  
 C:Superfamily: 6-phosphofructokinase, eukaryotic type; 6-phosphofructokinase 1 homology  
 C:Keywords: ATP; glycolysis; phosphoprotein; phosphotransferase  
 F:27-336/Domain: 6-phosphofructokinase 1 homology <6PPI>  
 F:413-699/Domain: 6-phosphofructokinase 1 homology <6PFI>

Query Match 79.5%; Score 35; DB 2; Length 791;  
 Best Local Similarity 85.7%; Pred. No. 4e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PSSPPE 8  
 DB 240 PSSPPE 246

RESULT 24  
 G75459  
 DNA-directed RNA polymerase, beta subunit - Deinococcus radiodurans (strain R1)  
 C:Species: Deinococcus radiodurans  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
 C:Accession: G75459  
 R:White, O.; Eilen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
 S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma  
 Science 286, 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; PMID:20036896; PMID:10567266  
 A:Accession: G75459  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1179 <WH1>  
 A:Cross-references: GB:AE001944; GB:AE000513; NID:G6458634; PIDN:AAF10490.1; PID:G645863  
 A:Experimental source: strain R1  
 C:Genetic:  
 A:Gene: DR0912  
 A:Map position: 1  
 C:Superfamily: DNA-directed RNA polymerase beta chain

Query Match 79.5%; Score 35; DB 2; Length 1179;  
 Best Local Similarity 85.7%; Pred. No. 6e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KPSPPPE 7  
 DB 140 KPSPPPE 146

RESULT 25  
 T12737  
 tail protein - Methanobacterium phage psiM2  
 N:Alternate names: protein 21  
 C:Species: Methanobacterium phage psiM2  
 C>Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 24-Sep-1999  
 C:Accession: T12737  
 R:Pieter, P.; Wasserfallen, A.; Stettler, R.; Leisinger, T.  
 submitted to the EMBL Data Library, May 1998  
 A:Description: Archaeophage PsiM2 complete genomic DNA.  
 A:Reference number: Z17578  
 A:Accession: T12737  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1186 <PFI>  
 A:Cross-references: EMBL:AF065411; NID:G3249585; PID:G3249606; PIDN:AACT27060.1  
 A:Experimental source: host Methanobacterium thermoautotrophicum strain Marburg

Query Match 79.5%; Score 35; DB 2; Length 1186;  
 Best Local Similarity 85.7%; Pred. No. 6.1e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PSSPPE 8  
 DB 1065 PSSPPE 1071

RESULT 26  
 D3B821  
 DNA-directed DNA polymerase (EC 2.7.7.7) - human herpesvirus 2 (strain 186)  
 C:Species: human herpesvirus 2  
 C>Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 27-Oct-2003  
 C:Accession: A27315  
 R:Tsurumi, T.; Maeno, K.; Nishiyama, Y.  
 Gene 52, 129-137, 1987  
 A:Title: Nucleotide sequence of the DNA polymerase gene of herpes simplex virus type 2 at  
 A:Reference number: A27315; PMID:87277385; PMID:3038677  
 A:Accession: A27315  
 A:Molecule type: DNA  
 A:Residues: 1-1240 <TSU>  
 A:Cross-references: GB:M16321; NID:G330291; PIDN:AAA5853.1; PID:G330292  
 C:Superfamily: DNA polymerase  
 C:Keywords: DNA binding; DNA replication; nucleotidyltransferase

Query Match 79.5%; Score 35; DB 1; Length 1240;  
 Best Local Similarity 85.7%; Pred. No. 6.3e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PSSPPE 8  
 DB 798 PSSPPE 804

RESULT 27  
 D54689  
 protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type delta, splice form D precursor  
 N:Alternate names: MPTP delta type D  
 C:Contents: protein tyrosine phosphatase, receptor type delta, splice form A  
 C:Species: Mus musculus (house mouse)  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: D54689; A54689  
 R:Mizuno, K.; Haegawa, K.; Katagiri, T.; Ogimoto, M.; Ichikawa, T.; Yakura, H.  
 Mol. Cell. Biol. 13, 5513-5523, 1993  
 A:Title: MPTP delta, a putative murine homolog of HPTP delta, is expressed in specialized  
 A:Reference number: A54689; PMID:93360986; PMID:8355697  
 A:Accession: D54689  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1691 <MI2>  
 A:Experimental source: brain  
 A:Note: sequence inconsistent with nucleotide translation  
 A:Accession: A54689  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-398,799-1691 <MI2>  
 A:Experimental source: brain  
 A:Note: sequence inconsistent with nucleotide translation  
 A:Note: sequence extracted from NCBI backbone (NCBIN:136522; NCBI:136524)  
 C:Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology; j  
 OGY  
 C:Keywords: alternative splicing; glycoprotein; phosphoprotein; phosphoric monoester hyd;  
 F:42-95/Domain: immunoglobulin homology <IMW3>  
 F:114-196/Domain: fibronectin type III repeat homology <FN3A>  
 F:1075-1691/Domain: leukocyte common antigen cytosolic domain homology <LAC>  
 F:1449-1671/Domain: protein-tyrosine-phosphatase homology <PTP2>  
 F:1333/Active site: Cys (phosphocysteine intermediate) #status predicted  
 F:1623/Active site: Cys (phosphocysteine intermediate) #status predicted  
 F:1629/Binding site: Cys (phosphocysteine intermediate) #status predicted  
 F:1629/Binding site: substrate phosphate (Arg) #status predicted

Query Match 79.5%; Score 35; DB 1; Length 1691;  
 Best Local Similarity 62.5%; Pred. No. 8,7e+02;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
 DB 399 KPSAPPOD 406

## RESULT 28

S46217  
 A:protein-tyrosine-phosphatase (EC 3.1.3.48) type sigma precursor - rat  
 N:Alternate names: leukocyte common antigen-related phosphatase  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 07-May-1995 #sequence revision 03-Nov-1995 #text\_change 23-Jul-1999  
 C:Accession: S46217; S51174; A49104  
 R:Zhang, W.R.; Hashimoto, N.; Ahmad, F.; Ding, W.; Goldstein, B.J.  
 Biochem. J. 302, 39-47, 1994  
 A:Title: Molecular cloning and expression of a unique receptor-like protein-tyrosine-ph  
 A:Reference number: S46216; MUID:94347119; PMID:8068021  
 A:Accession: S46217  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-1863 <ZHA>  
 A:Cross-references: EMBL:L11587  
 R:Goldstein, B.J.  
 Submitted to the EMBL Data Library, February 1993  
 A:Reference number: S51174  
 A:Accession: S51174  
 A:Molecule type: mRNA  
 A:Residues: 1-1788, 'G', 1790-1863 <GOL>  
 A:Cross-references: EMBL:L11587; NID:9205134; PID:AC37656.1; PID:9205135  
 R:Yan, H.; Grossman, A.; Wang, H.; D'Eustachio, P.; Mossie, K.; Musacchio, J.M.; Silvent  
 J. Biol. Chem. 268, 24880-24886, 1993  
 A:Title: A novel receptor tyrosine phosphatase-sigma that is highly expressed in the net  
 A:Reference number: A49104; MUID:94043351; PMID:8227050  
 A:Accession: A49104  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Residues: 1-596, 'R', 598-603, 'I', 967-1788, 'G', 1790-1863 <YAN>  
 A:Experimental source: brain  
 A>Note: sequence extracted from NCBI backbone (NCBIP:139669)  
 C:Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology;  
 ogy  
 C:Keywords: alternative splicing; glycoprotein; phosphoprotein; phosphoric monoester hyd  
 F:1-26/Domain: signal sequence #status predicted <SIG>  
 F:27-1863/Product: protein-tyrosine-phosphatase #status predicted <MAT>  
 F:149-209/Domain: immunoglobulin homology <IMM1>  
 F:246-300/Domain: immunoglobulin homology <IMM2>  
 F:318-400/Domain: fibronectin type III repeat homology <FN3A>  
 F:413-499/Domain: fibronectin type III repeat homology <FN3B>  
 F:511-592/Domain: fibronectin type III repeat homology <FN3C>  
 F:1244-1863/Domain: leukocyte common antigen cytosolic domain homology <LAC>  
 F:1331-1552/Domain: protein-tyrosine-phosphatase homology <PRP1>  
 F:1504/Active site: Cys (phosphocysteine intermediate) #status predicted  
 F:1510/Binding site: substrate phosphate (Arg) #status predicted  
 F:1795/Active site: Cys (phosphocysteine intermediate) #status predicted  
 F:1801/Binding site: substrate phosphate (Arg) #status predicted

Query Match 79.5%; Score 35; DB 2; Length 1863;  
 Best Local Similarity 62.5%; Pred. No. 9,6e+02;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
 DB 604 KPSAPPOD 611

## RESULT 29

C54689  
 A:protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type delta, splice form B precursor  
 N:Alternate names: MPTP delta type B/C

N:Contains: protein tyrosine phosphatase, receptor type delta, splice form C  
 C:Species: Mus musculus (house mouse)  
 C>Date: 25-Apr-1995 #sequence revision 19-May-1995 #text\_change 12-Feb-1999  
 C:Accession: C54689; B54689  
 R:Mizuno, K.; Hasegawa, K.; Katagiri, T.; Ogimoto, M.; Ichikawa, T.; Yakura, H.  
 Mol. Cell. Biol. 13, 5513-5523, 1993  
 A:Title: MPTP delta, a putative murine homology of HPTP delta, is expressed in specialized  
 A:Reference number: A54689; MUID:93360986; PMID:8355697  
 A:Accession: C54689  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1894 <MTZ>  
 A:Experimental source: brain; splice form B  
 A>Note: sequence inconsistent with nucleotide translation  
 A>Note: sequence extracted from NCBI backbone (NCBIN:137486, NCBIP:137487)  
 A:Accession: B54689  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-352, 'H', 354-535, 'S', 537-601, 1002-1894 <MT2>  
 A:Experimental source: brain; splice form C  
 A>Note: sequence inconsistent with nucleotide translation  
 A>Note: sequence extracted from NCBI backbone (NCBIP:136530)  
 C:Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology;  
 ogy

C:Keywords: alternative splicing; glycoprotein; phosphoprotein; phosphoric monoester hyd  
 F:45-107/Domain: immunoglobulin homology <IMM1>  
 F:245-299/Domain: immunoglobulin homology <IMM2>  
 F:317-399/Domain: fibronectin type III repeat homology <FN3A>  
 F:1278-1894/Domain: leukocyte common antigen cytosolic domain homology <LAC>  
 F:1652-1874/Domain: protein-tyrosine-phosphatase homology <PRP2>  
 F:1536/Active site: Cys (phosphocysteine intermediate) #status predicted  
 F:1542/Binding site: substrate phosphate (Arg) #status predicted  
 F:1826/Active site: Cys (phosphocysteine intermediate) #status predicted  
 F:1832/Binding site: substrate phosphate (Arg) #status predicted

Query Match 79.5%; Score 35; DB 2; Length 1894;  
 Best Local Similarity 62.5%; Pred. No. 9,7e+02;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPE 8  
 DB 602 KPSAPPOD 609

## RESULT 30

S50893  
 A:protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type sigma precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 01-Aug-1995 #sequence revision 01-Sep-1995 #text\_change 21-Jan-2000  
 C:Accession: S50893; S40281  
 R:Wagner, J.; Boerboom, D.; Tremblay, M.L.  
 Eur. J. Biochem. 226, 773-782, 1994  
 A:Title: Molecular cloning and tissue-specific RNA processing of a murine receptor-type I  
 A:Reference number: S50893; MUID:95112841; PMID:7529177  
 A:Accession: S50893  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1907 <WAG>  
 A:Cross-references: EMBL:X62288; NID:9587483; PID:CAA57732.1; PID:9587484  
 R:Hendriks, W.; Brugman, C.; Zeeuwen, P.; Schepens, J.; Wieringa, B.  
 Submitted to the EMBL Data Library, June 1993  
 A:Description: Assessment of the expression levels of murine protein-tyrosine phosphatase  
 A:Reference number: S40280  
 A:Accession: S40281  
 A:Molecule type: mRNA  
 A:Residues: 1441-1501, 'E', 1503-1546 <HEN>  
 A:Cross-references: EMBL:Z23050; NID:9438137; PID:CAA80585.1; PID:9438138  
 C:Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology;  
 ogy  
 C:Keywords: glycoprotein; phosphoprotein; phosphoric monoester hydrolase; transmembrane p  
 F:149-209/Domain: immunoglobulin homology <IMM1>  
 F:246-300/Domain: immunoglobulin homology <IMM2>  
 F:413-506/Domain: fibronectin type III repeat homology <FN3A>

F:1288-1907/Domain: leukocyte common antigen cytosolic domain homology <LAC>  
F:1375-1596/Domain: protein-tyrosine-phosphatase homology <PTP>  
F:1664-1887/Domain: protein-tyrosine-phosphatase homology <PTP2>  
F:1546/Active site: Cys (phosphotyrosine intermediate) #status predicted  
F:1554/Binding site: substrate phosphate (Arg) #status predicted  
F:1839/Active site: Cys (phosphotyrosine intermediate) #status predicted  
F:1845/Binding site: substrate phosphate (Arg) #status predicted

Query Match 79.5%; Score 35; DB 2; Length 1907;  
Best Local Similarity 62.5%; Pred. No. 9.8e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSPPEP 8  
Db 604 KFSAPPD 611

## RESULT 31

A56178  
protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type delta precursor - human  
N/Alternate names: protein-tyrosine-phosphatase BPTP-2  
C/Species: Homo sapiens (man)  
C/Date: 03-Oct-1995 #sequence revision 03-Oct-1995 #text\_change 21-Jan-2000  
C/Accession: A56178; S12052; B44929  
R/Pulido, R.; Krueger, N.X.; Serra-Pages, C.; Saito, H.; Streuli, M.  
J. Biol. Chem. 270, 6722-6728, 1995  
A/Title: Molecular characterization of the human transmembrane protein-tyrosine phosphatase delta isoforms.  
A/Reference number: A56178; PMID:95204468; PMID:7896816  
A/Accession: A56178  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-1912 <PU>  
A/Cross-references: GB:138929; NID:9755652; PIDN:AA041749.1; PID:9755653  
R/Krueger, N.X.; Streuli, M.; Saito, H.  
EMBO J. 9, 3241-3252, 1990  
A/Title: Structural diversity and evolution of human receptor-like protein tyrosine phosphatases  
A/Reference number: S12049; PMID:91006018; PMID:2170109  
A/Accession: S12052  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 390-1912 <KR>  
A/Cross-references: GB:X54133; NID:935789; PIDN:CA38068.1; PID:935790  
A/Note: the sequence from Fig. 5B is inconsistent with that from Fig. 5A in having 568-T  
R/Adachi, M.; Sekiya, M.; Arimura, Y.; Takekawa, M.; Iton, F.; Hinda, Y.; Imai, K.; Yac  
Cancer Res. 52, 737-740, 1992  
A/Title: Protein-tyrosine phosphatase expression in pre-B cell NALM-6.  
A/Reference number: A44929; PMID:92119637; PMID:1370651  
A/Accession: B44929  
A/Molecule type: mRNA  
A/Residues: 1756-1804, 'C', 1806-1845 <AD>  
A/Cross-references: GB:S78086; NID:9243545; PIDN:AAB21147.1; PID:9243546  
A/Experimental source: pre-B cell NALM-6  
A/Note: sequence extracted from NCBI backbone (NCBI:78086, NCBI:78087)  
A/Note: the authors did not report the entire codon for residue 90  
C/Genetic: GDB:PTPRD  
A/Genes: GDB:131384; OMIM:601598  
A/Cross-references: GDB:924-9p24  
A/Map position: 9p24-9p24  
C/Suprafamily: leukocyte antigen-related protein; fibronectin type III repeat homology;  
Og  
C/Keyword: glycoprotein; phosphoprotein; phosphoric monoester hydrolase; transmembrane  
F:138-100/Domain: immunoglobulin homology <IMW1>  
F:140-209/Domain: immunoglobulin homology <IMW2>  
F:250-304/Domain: immunoglobulin homology <IMW3>  
F:111-811/Domain: fibronectin type III repeat homology <3FR>  
F:1293-1912/Domain: leukocyte common antigen cytosolic domain homology <3FR>  
F:1659-1892/Domain: protein-tyrosine-phosphatase homology <PTP2>  
F:1553/Active site: Cys (phosphotyrosine intermediate) #status predicted  
F:1559/Binding site: substrate phosphate (Arg) #status predicted  
F:1844/Active site: Cys (phosphotyrosine intermediate) #status predicted  
F:1850/Binding site: substrate phosphate (Arg) #status predicted

Query Match 79.5%; Score 35; DB 2; Length 1912;  
Best Local Similarity 62.5%; Pred. No. 9.8e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSPPEP 8  
Db 608 KFSAPPD 615

## RESULT 32

FNBO

fibronectin - bovine  
C/Species: Bos primigenius taurus (cattle)  
C/Date: 31-Dec-1988 #sequence revision 31-Dec-1988 #text\_change 20-Oct-2000  
C/Accession: A26452; B21165; A23292  
R/Skorstenggaard, K.; Jensen, M.S.; Sahl, P.; Petersen, T.E.; Magnusson, S.  
Eur. J. Biochem. 161, 441-453, 1986  
A/Title: Complete primary structure of bovine plasma fibronectin.  
A/Reference number: A26452; PMID:87054047; PMID:3780752  
A/Accession: A26452  
A/Molecule type: protein  
A/Residues: 1-2265 <SKO>  
R/Kornblith, A.R.; Vibe-Pedersen, K.; Baralle, F.E.  
Proc. Natl. Acad. Sci. U.S.A. 80, 3218-3222, 1983  
A/Title: Isolation and characterization of cDNA clones for human and bovine fibronectins.  
A/Reference number: A21165; PMID:83221567; PMID:6304699  
A/Accession: B21165  
A/Molecule type: mRNA  
A/Residues: 2170-2265 <KOR>  
A/Cross-references: GB:K00800; NID:916305; PIDN:AAA0521.2; PID:95713323  
R/Petersen, T.E.; Thorgersten, H.C.; Skorstenggaard, K.; Vibe-Pedersen, K.; Sahl, P.; Sotter  
Proc. Natl. Acad. Sci. U.S.A. 80, 137-141, 1983  
A/Title: Partial primary structure of bovine plasma fibronectin: three types of internal  
A/Reference number: A23292; PMID:83117805; PMID:6218503  
A/Accession: A23292  
A/Molecule type: protein  
A/Residues: 1-16, 'C', 18-20, 'S', 22-43; 447-463; 1367-1517; 1567-1673; 2062-2176, 'N', 2178-2265  
C/Comment: Cys-1201 and Cys-2015 have free sulfhydryl groups.  
C/Comment: The plasma fibronectin molecule consists of two chains, which are connected by  
C/Comment: Fibronectin bind cell surfaces and various compounds including collagen, fibr  
aling, and maintenance of cell shape.  
C/Comment: Plasma fibronectin is synthesized by hepatocytes.  
C/Suprafamily: fibronectin; fibronectin type I repeat homology; fibronectin type II repe  
C/Keyword: acute phase; alternative splicing; collagen binding; duplication; extracellular  
F:21-241/Domain: fibronectin type I repeat homology <1F1>  
F:21-55/Domain: fibronectin type I repeat homology <1F2>  
F:66-104/Domain: fibronectin type I repeat homology <1F3>  
F:110-148/Domain: fibronectin type I repeat homology <1F4>  
F:155-194/Domain: fibronectin type I repeat homology <1F5>  
F:200-239/Domain: fibronectin type I repeat homology <1F6>  
F:277-577/Domain: collagen binding <CBR>  
F:277-311/Domain: fibronectin type I repeat homology <2F1>  
F:329-370/Domain: fibronectin type II repeat homology <2F2>  
F:389-430/Domain: fibronectin type II repeat homology <2F3>  
F:439-477/Domain: fibronectin type I repeat homology <1F7>  
F:487-524/Domain: fibronectin type I repeat homology <1F8>  
F:530-568/Domain: fibronectin type I repeat homology <1F9>  
F:578-661/Domain: fibronectin type III repeat homology <FN3A>  
F:688-770/Domain: fibronectin type III repeat homology <FN3B>  
F:779-860/Domain: fibronectin type III repeat homology <FN3C>  
F:875-957/Domain: fibronectin type III repeat homology <FN3D>  
F:965-1046/Domain: fibronectin type III repeat homology <FN3E>  
F:1055-1134/Domain: fibronectin type III repeat homology <FN3F>  
F:1142-1227/Domain: fibronectin type III repeat homology <FN3G>  
F:1235-1318/Domain: fibronectin type III repeat homology <FN3H>  
F:1326-1404/Domain: fibronectin type III repeat homology <FN3I>  
F:1410-1517/Domain: cell attachment <CAD>  
F:1416-1502/Domain: fibronectin type III repeat homology <FN3J>  
F:1493-1485/Region: cell attachment (R-G-D) motif  
F:1510-1592/Domain: fibronectin type III repeat homology <FN3K>  
F:1600-1870/Domain: heparin binding <HB2>  
F:1600-1682/Domain: fibronectin type III repeat homology <FN3L>  
F:1692-1773/Domain: fibronectin type III repeat homology <FN3M>

F:1181-1863/Domain: fibronectin type III repeat homology <FN3N>  
 F:1970-1972/Region: cell attachment (R-G-D) motif  
 F:1982-2062/Domain: fibronectin type III repeat homology <FN3O>  
 F:1985-2216/Domain: fibrin binding <FB2>  
 F:2085-2124/Domain: fibronectin type I repeat homology <FI10>  
 F:2130-2167/Domain: fibronectin type I repeat homology <FI1>  
 F:2174-2209/Domain: fibronectin type I repeat homology <FI12>  
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
 F:3/Cross-link: isopeptide (Gln) (interchain to Lys N6-amino of fibrin) #status expiring  
 F:21-47,45-56,66-94,92-104,110-138,136-148,155-184,182-194,200-229,227-239,277-304,302-3  
 7,2155-2167,2174-2200,2198-2209/Disulfide bonds: #status predicted  
 F:399,497,511,846,976,1213,1987/Binding site: carbonylrate (Asn) (covalent) #status expe  
 F:1205,1692/Binding site: carbonylrate (Asn) (covalent) #status absent  
 F:1943,1944/Binding site: carbonylrate (Thr) (covalent) #status experimental  
 F:2246/Disulfide bonds: interchain (to 2250) #status predicted  
 F:2250/Disulfide bonds: interchain (to 2246) #status predicted  
 F:2263/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 79.5%; Score 35; DB 1; Length 2265;  
 Best Local Similarity 75.0%; Pred. No. 1.2e+03;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KPSPPPE 8  
 Db 1818 KPSPPPE 1825

RESULT 33  
 fibronectin precursor [validated] - human  
 N/Alternate names: fibronectin splice form ED-A  
 C/Species: Homo sapiens (hmn)  
 C/Date: 27-Nov-1985 #sequence, revision 31-Mar-1993 #text, change 08-Dec-2000  
 C/Accession: A26460; A26284; S03917; A24476; A24476; A91008; A93529; A21011; A90495; A22  
 R/Dear: D.C.; Bowler, C.L.; Bourgeois, S.  
 Proc. Natl. Acad. Sci. U.S.A. 84, 1876-1880, 1987  
 A/Title: Cloning and analysis of the promoter region of the human fibronectin gene.  
 A/Reference number: A26460; MUID:87175578; PMID:3031656  
 A/Accession: A26460  
 A/Molecule type: DNA  
 A/Residues: 1-49 <DEA>  
 A/Cross-references: GB:M15801; NID:g182686; PIDN:AAA53376.1; PID:g553293  
 R/Oldberg, A.; Ruoslahti, E.  
 J. Biol. Chem. 261, 2113-2116, 1986  
 A/Title: Evolution of the fibronectin gene.  
 A/Reference number: A26284; MUID:86111901; PMID:3003095  
 A/Accession: A26284  
 A/Molecule type: DNA  
 A/Residues: 1447-1540 <OLD>  
 A/Cross-references: GB:M12549; NID:g182688  
 A/Note: the authors translated the codon TTC for residue 1494 as Glu  
 R/Paoletti, G.; Hencicliffe, C.; Sebastio, G.; Baralle, F.E.  
 Nucleic Acids Res. 16, 3545-3557, 1988  
 A/Title: Sequence analysis and in vivo expression show that alternative splicing of ED-E  
 A/Reference number: S00848; MUID:88233940; PMID:3375063  
 A/Accession: S03917  
 A/Molecule type: DNA  
 A/Residues: 1594-1767, 'V', 1769-1783 <PAO>  
 A/Cross-references: EMBL:X07718; NID:g31402  
 A/Note: the authors translated the codon AAC for residue 1631 as Asp  
 R/Vibe-Pedersen, K.; Magnusson, S.; Baralle, F.E.  
 FEBS Lett. 207, 287-291, 1986  
 A/Title: Donor and acceptor splice signals within an exon of the human fibronectin gene:  
 A/Reference number: A24854; MUID:87030929; PMID:3770201  
 A/Accession: A24854  
 A/Molecule type: DNA  
 A/Residues: 1992-2147 <VIB>  
 A/Cross-references: GB:X04530; NID:g31436  
 R/Gutman, A.; Yamada, K.M.; Kornblith, A.  
 FEBS Lett. 207, 145-148, 1986  
 A/Title: Human fibronectin is synthesized as a pre-propolypeptide.  
 A/Reference number: A24476; MUID:87030890; PMID:33770189  
 A/Accession: A24476

A/Status: not compared with conceptual translation  
 A/Molecule type: mRNA  
 A/Residues: 1-14, 'Q', 16-38 <GUT>  
 R/Kornblith, A.R.; Umezawa, K.; Vibe-Pedersen, K.; Baralle, F.E.  
 EMBL J. 4, 1755-1759, 1985  
 A/Title: Primary structure of human fibronectin: differential splicing may generate at 1e  
 A/Reference number: A91008; MUID:85284965; PMID:2592939  
 A/Accession: A91008  
 A/Molecule type: mRNA  
 A/Status: nucleic acid sequence not shown  
 A/Residues: 32-144, 1346-2080, 2112-2386 <KOR>  
 A/Cross-references: GB:X02761  
 R/Kornblith, A.R.; Vibe-Pedersen, K.; Baralle, F.E.  
 Nucleic Acids Res. 12, 5851-5868, 1984  
 A/Title: Human fibronectin: cell specific alternative mRNA splicing generates polypeptide  
 A/Reference number: A93529; MUID:84272258; PMID:6462919  
 A/Accession: A93529  
 A/Molecule type: mRNA  
 A/Residues: 973-2080, 2112-2386 <KO2>  
 A/Cross-references: GB:X00739  
 R/Oldberg, A.; Linney, E.; Ruoslahti, E.  
 J. Biol. Chem. 258, 10193-10196, 1983  
 A/Title: Molecular cloning and nucleotide sequence of a cDNA clone coding for the cell at  
 A/Reference number: A21011; MUID:83250929; PMID:6688418  
 A/Accession: A21011  
 A/Molecule type: mRNA  
 A/Residues: 1434-1537 <OL2>  
 A/Cross-references: GB:K00055; NID:g182680; PIDN:AAA52459.1; PID:g182683  
 R/Bernard, M.P.; Kolbe, M.; Weil, D.; Chu, M.L.  
 Biochemistry 24, 2698-2704, 1985  
 A/Title: Human cellular fibronectin: comparison of the carboxyl-terminal portion with rat  
 A/Reference number: A90495; MUID:85280409; PMID:2592573  
 A/Accession: A90495  
 A/Molecule type: mRNA  
 A/Residues: 1594-2386 <BER>  
 A/Cross-references: GB:M10905; NID:g182696; PIDN:AAA52462.1; PID:g182697  
 R/Umezawa, K.; Kornblith, A.R.; Baralle, F.E.  
 FEBS Lett. 186, 31-34, 1985  
 A/Title: Isolation and characterization of cDNA clones for human liver fibronectin.  
 A/Reference number: A22245; MUID:85231203; PMID:2989004  
 A/Accession: A22245  
 A/Molecule type: mRNA  
 A/Residues: 1948-2067 <UMB>  
 A/Cross-references: GB:M27589; NID:g182705; PIDN:AAA52465.1; PID:g182706  
 A/Accession: B22245  
 A/Molecule type: mRNA  
 A/Residues: 1975-1991, 2017-2039 <UM2>  
 A/Cross-references: GB:M27590  
 R/Sekiguchi, K.; Klob, A.M.; Kuyachi, K.; Yoshitake, S.; Hakomori, S.  
 Biochemistry 25, 4936-4941, 1986  
 A/Title: Human liver fibronectin complementary DNAs: identification of two different mess  
 A/Reference number: 152394; MUID:87026578; PMID:3021206  
 A/Accession: 152394  
 A/Status: preliminary; translated from GB/EMBL/DBD  
 A/Molecule type: mRNA  
 A/Residues: 1978-1990, 2016-2018, 'N', 2020-2081, 2113-2127 <SEK>  
 A/Cross-references: GB:M14060; NID:g182701; PIDN:AAA52464.1; PID:g182704  
 R/Kornblith, A.R.; Vibe-Pedersen, K.; Baralle, F.E.  
 Proc. Natl. Acad. Sci. U.S.A. 80, 3218-3222, 1983  
 A/Title: Isolation and characterization of cDNA clones for human and bovine fibronectins  
 A/Reference number: A21165; MUID:83221567; PMID:6504699  
 A/Accession: A21165  
 A/Molecule type: mRNA  
 A/Residues: 2291-2386 <KO3>  
 A/Cross-references: GB:K00739; NID:g182681; PIDN:AAA52460.1; PID:g182684  
 R/Garcia-Pardo, A.; Pearlstein, E.; Frangione, B.  
 J. Biol. Chem. 258, 12670-12674, 1983  
 A/Title: Primary structure of human plasma fibronectin.  
 A/Reference number: A92398; MUID:84032463; PMID:6630202  
 A/Accession: A92398  
 A/Molecule type: protein  
 A/Residues: 32-47, 'C', 49-51, 'S', 53-72, 'A', 74-290 <GAR1>  
 R/Garcia-Pardo, A.; Gold, L.I.

Arch. Biochem. Biophys. 304, 181-188, 1993  
 A>Title: Further characterization of the binding of fibronectin to gelatin reveals the  
 A:Reference number: S34791; PMID:93312001; PMID:8323285  
 A:Accession: S34791  
 A:Molecule type: protein  
 A:Residues: 291-300;551-560 <GAR2>  
 R:Griffin, C.A.; Calaycay, J.; Shively, J.E.; Smith, R.L.  
 Thromb. Res. 43, 469-477, 1986  
 A>Title: Two plasma fibronectin fragments with different gelatin-binding properties.  
 A:Reference number: A60904; PMID:87019725; PMID:3532418  
 A:Accession: A60904  
 A:Molecule type: protein  
 A:Residues: 293-301 <GRI>  
 R:Calaycay, J.; Pandey, H.; Lee, T.; Borsi, L.; Sirt, A.; Shively, J.E.; Zardi, L.  
 J. Biol. Chem. 260, 12136-12141, 1985  
 A>Title: Primary structure of a DNA- and heparin-binding domain (domain III) in human p  
 A:Reference number: A23901; PMID:86008277; PMID:3900070  
 A:Accession: A23901  
 A:Molecule type: protein  
 A:Residues: 616-677 'Q', 679-703 'PT' <CAL>  
 R:Pierichbacher, M.D.; Ruoslahti, E.; Sundelin, J.; Lind, P.; Peterson, P.A.  
 J. Biol. Chem. 257, 9593-9597, 1982  
 A>Title: The cell attachment domain of fibronectin. Determination of the primary structu  
 A:Reference number: A92386; PMID:82265604; PMID:7050098  
 A:Accession: A92386  
 A:Molecule type: protein  
 A:Residues: 1441-1548 <PIE>  
 A:Note: residues 1524-1527 are responsible for the cell-binding activity  
 R:Garcia-Pardo, A.; Roostagno, A.; Frangione, B.  
 Biochem. J. 241, 923-928, 1987  
 A>Title: Primary structure of human plasma fibronectin. Characterization of a 38 kDa dom  
 A:Reference number: A32517; PMID:87241275; PMID:3593220  
 A:Accession: A32517  
 A:Molecule type: protein  
 A:Residues: 1589-1630 'T', 1722-2058 <GAR3>  
 R:Tresselt, T.; McCarthy, J.B.; Calaycay, J.; Lee, T.D.; Legesse, K.; Shively, J.E.; Pand  
 Blochem. J. 274, 731-738, 1991  
 A>Title: Human plasma fibronectin. Demonstration of structural differences between the A  
 A:Reference number: S14357; PMID:91190085; PMID:2012601  
 A:Accession: S14357  
 A:Molecule type: protein  
 A:Residues: 1614-1630 'T', 1722-2081, 2113-2244 <TRR>  
 R:Garcia-Pardo, A.; Pearlstein, E.; Frangione, B.  
 J. Biol. Chem. 260, 10320-10325, 1985  
 A>Title: Primary structure of human plasma fibronectin. Characterization of a 31,000-dal  
 A:Reference number: A23891; PMID:85261459; PMID:4019516  
 A:Accession: A23891  
 A:Molecule type: protein  
 A:Residues: 2071-2080;2112-2356 <GAR4>  
 C:Comment: The extra domain and connecting strand 3 are subject to developmental and tis  
 C:Comment: The cellular and plasma fibronectins are high molecular weight glycoproteins,  
 ation, and transformation.  
 C:Genetic:  
 A:Gene: GDB:FNI  
 A:Cross-references: GDB:119135; OMIM:135600  
 A:Map position: 2q34-2q34  
 A:Residues: 49/3; 1266/1; 1357/1; 1447/1; 1487/1; 1541/1; 1631/1; 1721/1; 1991/1; 2145/1  
 C:Superfamily: fibronectin; fibronectin type I repeat homology; fibronectin type II rep  
 C:Keywords: acute phase; alternative splicing; cell adhesion; collagen binding; duplicat  
 F:1-26/Domain: signal sequence #status predicted <SIG>  
 F:127-131/Domain: propeptide #status predicted <PRO>  
 F:132-2386/Product: fibronectin #status experimental <MAT>  
 F:152-272/Domain: fibronectin type I repeat homology <FIB>  
 F:152-97/Domain: fibronectin type I repeat homology <FIB>  
 F:197-135/Domain: fibronectin type I repeat homology <FIB>  
 F:141-179/Domain: fibronectin type I repeat homology <FIB>  
 F:186-225/Domain: fibronectin type I repeat homology <FIB>  
 F:231-270/Domain: fibronectin type I repeat homology <FIB>  
 F:308-608/Domain: collagen binding <CBR>  
 F:308-342/Domain: fibronectin type I repeat homology <FIB>  
 F:360-401/Domain: fibronectin type II repeat homology <FIB>  
 F:420-461/Domain: fibronectin type II repeat homology <FIB>  
 F:470-508/Domain: fibronectin type I repeat homology <FIB>

F:518-555/Domain: fibronectin type I repeat homology <FIB>  
 F:561-599/Domain: fibronectin type I repeat homology <FIB>  
 F:609-692/Domain: fibronectin type III repeat homology <FIB>  
 F:616-706/Domain: heparin binding <HBP>  
 F:719-801/Domain: fibronectin type III repeat homology <FIB>  
 F:810-891/Domain: fibronectin type III repeat homology <FIB>  
 F:906-988/Domain: fibronectin type III repeat homology <FIB>  
 F:996-1077/Domain: fibronectin type III repeat homology <FIB>  
 F:1086-1164/Domain: fibronectin type III repeat homology <FIB>  
 F:1173-1258/Domain: fibronectin type III repeat homology <FIB>  
 F:1266-1349/Domain: fibronectin type III repeat homology <FIB>  
 Query Match 79.5%; Score 35; DB 1; Length 2386;  
 Best Local Similarity 75.0%; Pred. No. 12e+03;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 KESPPPE 8  
 Db 1939 KPCSPPE 1946  
 RESULT 34  
 S14428  
 fibronectin precursor - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 28-Oct-1994 #sequence revision 28-Oct-1994 #text change 20-Aug-1999  
 C:Accession: S14428; S12455; A22319; S46203; S00459; A27252; I59049  
 R:Hynes, R.O.  
 Submitted to the EMBL Data Library, July 1989  
 A:Reference number: S14428  
 A:Accession: S14428  
 A:Molecule type: mRNA  
 A:Residues: 1-2477 <HYN>  
 A:Cross-references: EMBL:X15906; NID:956163; PID:CA34020.1; PID:956164  
 R:Schwarzbauer, J.E.; Patel, R.S.; Fonda, D.; Hynes, R.O.  
 EMBL J. 6, 2573-2580, 1987  
 A>Title: Multiple sites of alternative splicing of the rat fibronectin gene transcript.  
 A:Reference number: S12455; PMID:88054951; PMID:2445560  
 A:Accession: S12455  
 A>Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 609-1810 'T', 1812-2283 <SCH>  
 A:Cross-references: EMBL:X15906  
 R:Tamkun, J.W.; Schwarzbauer, J.E.; Hynes, R.O.  
 Proc. Natl. Acad. Sci. U.S.A. 81, 5140-5144, 1984  
 A>Title: A single rat fibronectin gene generates three different mRNAs by alternative sp  
 A:Reference number: A22319; PMID:84298097; PMID:6089177  
 A:Accession: A22319  
 A:Molecule type: DNA  
 A:Residues: 2052-2237 <TAM>  
 R:Palckenberg, C.; Englund, J.J.; Thogersen, I.B.; Salvesen, G.; Akerstrom, B.  
 Blochem. J. 301, 745-751, 1994  
 A>Title: Isolation and characterization of fibronectin-alpha(1)-microglobulin complex in  
 A:Reference number: S46203; PMID:94330948; PMID:7519849  
 A:Accession: S46203  
 A>Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1183-1192; GLN, 1268, 'P', 1270-1271, 'D', 1273, 'CF', 1276, 'PY', 1385-1399 <PAL>  
 R:Patel, R.S.; Odermatt, E.; Schwarzbauer, J.E.; Hynes, R.O.  
 EMBL J. 6, 2565-2572, 1987  
 A>Title: Organization of the fibronectin gene provides evidence for exon shuffling during  
 A:Reference number: S00459; PMID:88054950; PMID:3119923  
 A:Accession: S00459  
 A:Molecule type: DNA  
 A:Residues: 1-139;2382-2477 <PAT>  
 A:Cross-references: EMBL:X05831  
 A:Note: the authors translated the codon CCG for residues 51 and 94 as Ala  
 R:Schwarzbauer, J.E.; Tamkun, J.W.; Lemischka, I.R.; Hynes, R.O.  
 Cell 35, 421-431, 1983  
 A>Title: Three different fibronectin mRNAs arise by alternative splicing within the codi  
 A:Reference number: A27252; PMID:84082067; PMID:6317187  
 A:Accession: A27252  
 A:Molecule type: mRNA



A/Residues: 1586-1720, 'T', 1722, 1813-2477 <SC2>  
 R/odermat. E.; Tamkun, J.W.; Hynes, R.O.  
 Proc. Natl. Acad. Sci. U.S.A. 82, 6571-6575, 1985  
 A/Title: Repeating modular structure of the fibronectin gene: Relationship to protein st  
 A/Reference number: 159049; MUID:6016741; PMID:3863113  
 A/Accession: 159049  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1722-1810 <RES>  
 A/Cross-references: GB:M11750; NID:g204164; PIDN:AAA1170.1; PID:g554437  
 C/Genetics:  
 A/Introns: 51/1, 94/1, 2416/3, 2454/3  
 C/Superfamily: fibronectin; fibronectin type I repeat homology; fibronectin type II repe  
 C/Keywords: alternative splicing; cell adhesion; collagen binding; disulfide bond; dupli  
 F/1-32/Domain: signal sequence #status predicted <SIG>  
 F/33-2477/Product: fibronectin #status predicted <MAT>  
 F/53-88/Domain: fibronectin type I repeat homology <1P1>  
 F/98-136/Domain: fibronectin type I repeat homology <1P2>  
 F/142-180/Domain: fibronectin type I repeat homology <1P3>  
 F/187-226/Domain: fibronectin type I repeat homology <1P4>  
 F/232-271/Domain: fibronectin type I repeat homology <1P5>  
 F/308-342/Domain: fibronectin type I repeat homology <1P6>  
 F/360-401/Domain: fibronectin type II repeat homology <2P1>  
 F/420-461/Domain: fibronectin type II repeat homology <2P2>  
 F/470-508/Domain: fibronectin type I repeat homology <1P7>  
 F/518-555/Domain: fibronectin type I repeat homology <1P8>  
 F/561-599/Domain: fibronectin type I repeat homology <1P9>  
 F/609-692/Domain: fibronectin type II repeat homology <FN3A>  
 F/718-800/Domain: fibronectin type III repeat homology <FN3B>  
 F/809-890/Domain: fibronectin type III repeat homology <FN3C>  
 F/905-987/Domain: fibronectin type III repeat homology <FN3D>  
 F/995-1076/Domain: fibronectin type III repeat homology <FN3E>  
 F/1085-1164/Domain: fibronectin type III repeat homology <FN3F>  
 F/1172-1257/Domain: fibronectin type III repeat homology <FN3G>  
 F/1265-1348/Domain: fibronectin type III repeat homology <FN3H>  
 F/1356-1439/Domain: fibronectin type III repeat homology <FN3I>  
 F/1447-1529/Domain: fibronectin type III repeat homology <FN3J>  
 F/1537-1619/Domain: fibronectin type III repeat homology <FN3K>  
 F/1614-1616/Region: cell attachment (R-G-D) motif  
 F/1631-1713/Domain: fibronectin type III repeat homology <FN3L>  
 F/1721-1803/Domain: fibronectin type III repeat homology <FN3M>  
 F/1811-1893/Domain: fibronectin type III repeat homology <FN3N>  
 F/1903-1984/Domain: fibronectin type III repeat homology <FN3O>  
 F/1992-2074/Domain: fibronectin type III repeat homology <FN3P>  
 F/2181-2183/Region: cell attachment (R-G-D) motif  
 F/2193-2273/Domain: fibronectin type III repeat homology <FN3Q>  
 F/2296-2335/Domain: fibronectin type I repeat homology <1P10>  
 F/2341-2378/Domain: fibronectin type I repeat homology <1P11>  
 F/2385-2420/Domain: fibronectin type I repeat homology <1P12>  
 F/53-79, 77-88, 98-126, 124-136, 142-170, 168-180, 187-216, 214-226, 232-261, 259-271, 308-335, 333  
 368, 2366-2378, 2385-2411, 2409-2420/Disulfide bonds: #status predicted  
 F/2458/Disulfide bonds: interchain (to 2462) #status predicted  
 F/2462/Disulfide bonds: interchain (to 2458) #status predicted

Query Match 79.5%; Score 35; DB 2; Length 2477;  
 Best Local Similarity 75.0%; Pred. No. 1.3e+03;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KPSPPEE 8  
 Db 2029 KPSPPEE 2036

RESULT 35  
 AG2522  
 hypothetical protein asl7359 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120a1  
 C/Species: Nostoc sp. PCC 7120  
 A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C/Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
 C/Accession: AG2522  
 R/Kanehko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi,  
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.  
 DNA Res. 8, 205-213, 2001

A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anat  
 A/Reference number: AB1807; MUID:21595285; PMID:11759840  
 A/Accession: AG2522  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-66 <KOR>  
 A/Cross-references: GB:BA000020; PIDN:BBB77117.1; PID:gl7134558; GSPDB:GN00180  
 A/Experimental source: strain PCC 7120  
 C/Genetics:  
 A/Gene: asl7359  
 A/Genome: plasmid

Query Match 77.3%; Score 34; DB 2; Length 66;  
 Best Local Similarity 62.5%; Pred. No. 47;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KPSPPEE 8  
 Db 30 KPSPPEE 37

RESULT 36  
 F83167  
 hypothetical protein PA3833 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C/Species: Pseudomonas aeruginosa  
 C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C/Accession: F83167  
 R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laribgy, K.; Lim,  
 .; Lof, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathog  
 A/Reference number: AB2950; MUID:20437337; PMID:10984043  
 A/Accession: F83167  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-122 <STO>  
 A/Cross-references: GB:AE004800; GB:AE004091; NID:g9949999; PIDN:AA607220.1; GSPDB:GN001;  
 A/Experimental source: strain PA01  
 C/Genetics:  
 A/Gene: PA3833

Query Match 77.3%; Score 34; DB 2; Length 122;  
 Best Local Similarity 62.5%; Pred. No. 88;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KPSPPEE 8  
 Db 5 KPSPPEE 12

RESULT 37  
 H82468  
 hypothetical protein VCA0370 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
 C/Species: Vibrio cholerae  
 C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
 C/Accession: H82468  
 R/Heldberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;  
 chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bas, S.; Qin, H.; Dragoi, I.; Sellers, P  
 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 406, 477-483, 2000  
 A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
 A/Reference number: AB2035; MUID:20406833; PMID:10952301  
 A/Accession: H82468  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-124 <HET>  
 A/Cross-references: GB:AE004373; GB:AE003853; NID:g9657758; PIDN:AAF66277.1; GSPDB:GN001;  
 A/Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C/Genetics:  
 A/Gene: VCA0370; VCA0369  
 A/Map position: 2



Query Match 77.3%; Score 34; DB 2; Length 124;  
Best Local Similarity 85.7%; Pred. No. 90;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KPSSPP 7  
DB 62 KPMSPP 68

RESULT 38  
E72759  
hypothetical protein APE0066 - Aeropyrum pernix (strain K1)  
C/Species: Aeropyrum pernix  
C/Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jun-2000  
C/Accession: E72759  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999  
A/Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr A/Reference number: A72450; MUID:99310339; PMID:10382966  
A/Accession: E72759  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-136 <RAM>  
A/Cross-references: DDBJ:AP000058; NID:G5103388; PIDN:BAA78975.1; PID:dl042751; PID:G510 C/Genetics:  
A/Experimental source: strain K1

A:gene: APE0066  
C:Superfamily: Aeropyrum pernix hypothetical protein APE0066

Query Match 77.3%; Score 34; DB 2; Length 136;  
Best Local Similarity 75.0%; Pred. No. 98;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KPSSPBE 8  
DB 35 KPSSPBE 42

RESULT 39  
E72505  
hypothetical protein APE2019 - Aeropyrum pernix (strain K1)  
C/Species: Aeropyrum pernix  
C/Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jun-2000  
C/Accession: E72505  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999  
A/Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr A/Reference number: A72450; MUID:99310339; PMID:10382966  
A/Accession: E72505  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-140 <RAM>  
A/Cross-references: DDBJ:AP000063; NID:G5105654; PIDN:BAA81029.1; PID:dl044815; PID:G510 C/Genetics:  
A/Experimental source: strain K1

A:gene: APE2019  
C:Superfamily: Aeropyrum pernix hypothetical protein APE2019

Query Match 77.3%; Score 34; DB 2; Length 140;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPP 6  
DB 86 KPSSPP 91

RESULT 40  
T31243  
hypothetical protein 915 - Sphingomonas aromaticivorans plasmid pNL1  
C/Species: Sphingomonas aromaticivorans

C/Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 09-Jun-2000  
C/Accession: T31243  
R:Romine, M.F.; Stillwell, L.C.; Wong, K.K.; Thureston, S.J.; Sisk, E.C.; Sengen, C.W.; Ge submitted to the EMBL Data Library, July 1998  
A/Description: Complete sequence of a 184 kb catabolic plasmid from Sphingomonas aromatic A/Reference number: Z20992  
A/Accession: T31243  
A/Status: preliminary; translated from GB/EMBL/DDBJ  
A/Molecule type: DNA  
A/Residues: 1-173 <ROM>  
A/Cross-references: EMBL:AF079317; NID:G3378261; PID:G3378383; PIDN:AAD03966.1 C/Genetics:  
A/Genome: plasmid pNL1  
C/Note: orf915  
C:Superfamily: Sphingomonas aromaticivorans hypothetical protein 915

Query Match 77.3%; Score 34; DB 2; Length 173;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPP 6  
DB 13 KPSSPP 18

RESULT 41  
H82155  
hypothetical protein VCI1788 VCA0370 [imported] - Vibrio cholerae (strain N16961 serogrou C/Species: Vibrio cholerae  
C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C/Accession: H82155; A82469  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bacs, S.; Qin, H.; Dragoi, I.; Sellers, P. l, R.R.; Mekalanos, M.V.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A/Title: DNA Sequence of Both chromosomes of the cholera pathogen Vibrio cholerae.  
A/Reference number: A82035; MUID:20406833; PMID:10952301  
A/Accession: H82155  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-231 <H81>  
A/Cross-references: GB:AE004256; GB:AE003852; NID:G9656310; PIDN:AAF94937.1; GSPDB:GN001 A/Experimental source: serogroup O1; strain N16961; biotype El Tor  
A/Accession: A82469  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-231 <H82>  
A/Cross-references: GB:AE004373; GB:AE003853; NID:G9657758; PIDN:AAF966276.1; GSPDB:GN001 A/Experimental source: serogroup O1; strain N16961; biotype El Tor  
C/Genetics: <GEN1>  
A/Gene: VCI1788  
A/Map position: 1  
A/Map position: 1  
C/Genetics: <GEN2>  
A/Gene: VCA0370  
A/Map position: 2

Query Match 77.3%; Score 34; DB 2; Length 231;  
Best Local Similarity 85.7%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KPSSPBE 7  
DB 169 KPMSPP 175

RESULT 42  
A60548  
synaptohysein - Pacific electric ray  
C/Species: Torpedo californica (Pacific electric ray)  
C/Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 04-Sep-1998  
C/Accession: A60548  
R:Cowan, D.; Linial, M.; Scheller, R.H.  
Brain Res. 509, 1-7, 1990

A>Title: Torpedo synaptophysin: evolution of a synaptic vesicle protein.  
 A/Reference number: A60548; MUID:90167450; PMID:2106365  
 A/Accession: A60548  
 A/Status: not compared with conceptual translation  
 A/Molecule type: mRNA  
 A/Residues: 1-301 <COM>  
 C/Superfamily: synaptophysin  
 C/Keywords: channel-forming protein; glycoprotein; hexamer; membrane protein; phosphoric

Query Match 77.3%; Score 34; DB 2; Length 301;  
 Best Local Similarity 75.0%; Pred. No. 2.2e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8  
 |||||  
 Db 229 KPGLPPEE 236

## RESULT 43

TJ3972  
 hypothetical protein DC2.3 - Caenorhabditis elegans  
 C/Species: Caenorhabditis elegans  
 C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C/Accession: TJ3972  
 R/Bauer, C.; Sammons, L.; Wohlman, P.; Chiapelli, B.  
 submitted to the EMBL Data Library, February 1999  
 A/Description: The sequence of C. elegans cosmid DC2.  
 A/Reference number: Z21447  
 A/Accession: TJ3972

A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-310 <BAN>  
 A/Cross-references: EMBL:AF125956; PIDN:AA14724.1; GSPDB:GN00023; CESP:DC2.3  
 A/Experimental source: strain Bristol N2; clone DC2  
 C/Genetics:

A/Map position: 5  
 A/Intons: 65/2; 130/1; 150/3; 184/2; 217/2; 264/3; 303/3

Query Match 77.3%; Score 34; DB 2; Length 310;  
 Best Local Similarity 62.5%; Pred. No. 2.3e+02;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8  
 |||||  
 Db 135 EPSAPPDE 142

## RESULT 44

G84900  
 G-box binding bZIP transcription factor [imported] - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
 C/Accession: G84900  
 R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanhaken, S.E.; Umayam, L.; Tallon, L.;  
 Eues, D.; Nierman, W.C.; White, O.; Eissen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A/Reference number: A84420; MUID:20083487; PMID:10617197

A/Accession: G84900  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-382 <STO>  
 A/Cross-references: GB:AE002093; NID:G3702322; PIDN:AA62879.1; GSPDB:GN00139  
 C/Genetics:  
 A/Map position: 2  
 A/Map position: 2  
 C/Superfamily: fava bean G-box-binding protein; fos/jun DNA-binding domain homology

Query Match 77.3%; Score 34; DB 2; Length 382;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPP 6  
 |||||  
 Db 16 KPSSPP 21

## RESULT 45

UC7868  
 triacylglycerol lipase (EC 3.1.1.3) - Kurtzmanomyces sp. I-11  
 N/Alternate names: lipase; triacylglycerol acylhydrolase; tributyrase; triglyceride lipase  
 C/Species: Kurtzmanomyces sp. I-11  
 C/Date: 09-Dec-2002 #sequence\_revision 09-Dec-2002 #text\_change 31-Mar-2003  
 C/Accession: UC7868; PC7198  
 R/Kakugawa, K.; Shobayashi, M.; Suzuki, O.; Miyakawa, T.  
 Biosci. Biotechnol. Biochem. 66, 1328-1336, 2002  
 A/Title: Cloning, characterization, and expression of cDNA encoding a lipase from Kurtzme  
 A/Reference number: UC7868; MUID:22152189; PMID:12162555

A/Accession: UC7868  
 A/Molecule type: mRNA  
 A/Residues: 1-458 <KAK>  
 A/Cross-references: DDBJ:AB073866  
 A/Accession: PC7198

A/Molecule type: protein  
 A/Residues: 27-56,133-152,188-207,301-317,340-357 <KA2>  
 C/Comment: This enzyme, which is a thermostable, acidophilic and position-non-specific ly  
 n of glycerides. Thus, this enzyme is useful as a catalyst in industry such as ester synt  
 C/Genetics:  
 A/Map position: 11p1  
 C/Keywords: carboxylic ester hydrolase

Query Match 77.3%; Score 34; DB 2; Length 458;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPP 6  
 |||||  
 Db 96 KPSSPP 101

## RESULT 46

A42401  
 macrophage elastase (EC 3.4.24.-) precursor - mouse  
 N/Alternate names: matrix metalloproteinase 12 (MMP12)  
 C/Species: Mus musculus (house mouse)  
 C/Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 22-Jun-1999  
 C/Accession: A42401  
 R/Shapiro, S.D.; Griffin, G.L.; Gilbert, D.J.; Jenkins, N.A.; Copeland, N.G.; Welgus, H.C.  
 J. Biol. Chem. 267, 4664-4671, 1992  
 A/Title: Molecular cloning, chromosomal localization, and bacterial expression of a murin  
 A/Reference number: A42401; MUID:92165826; PMID:1537850  
 A/Accession: A42401

A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-462 <SHA>  
 A/Cross-references: GB:M82831; NID:G199127; PIDN:AA35526.1; PID:G199128  
 C/Superfamily: Interstitial collagenase; hemopexin repeat homology; matrix metalloprotein  
 C/Keywords: hydrolase; metalloproteinase; zinc; zymogen  
 F:53-256/Domain: matrix metalloproteinase homology <MMP>  
 F:269-462/Domain: hemopexin repeat homology <PKN>  
 F:85,211,215,221/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status I  
 F:211,215,221/Binding site: zinc, catalytic (His) (active) #status predicted  
 F:212/Active site: Glu #status predicted

Query Match 77.3%; Score 34; DB 2; Length 462;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPP 6  
 |||||  
 Db 266 KPSSPP 271

## RESULT 47

## FOVWLV

gag polypeptide - human immunodeficiency virus type 1 (isolate LAV-1a)

N:Alternate names: assemblin; core polypeptide; gag precursor

N:Contains: capsid antigen core protein p24CA; core protein p1; core protein p6; matrix

C:Species: human immunodeficiency virus type 1, HIV-1

A:Note: host Homo sapiens (man)

C:Date: 17-May-1985 #sequence\_rev1501 17-May-1985 #text\_change 05-Dec-1998

C:Accession: A03946

R:Main-Hobson, S.; Sonigo, P.; Danos, O.; Cole, S.; Alizon, M.

Cell 40, 9-11, 1985

A:Title: Nucleotide sequence of the AIDS virus, LAV.

A:Reference number: A90866; PMID:8509333; PMID:2981635

A:Accession: A03946

A:Molecule type: DNA

A:Residues: 1-500 <WAI>

C:Genetics:

A:Gene: gag

C:Superfamily: AIDS-related virus gag polypeptide

C:Keywords: AIDS; blocked amino end; core protein; immunodeficiency; lipoprotein; myrist

F:2-500/Product: gag precursor (assemblin) #status predicted <GAG>

F:2-132/Product: matrix antigen core protein p17MA #status predicted <P17>

F:20-32/Region: nuclear location signal

F:110-114/Region: nuclear location signal

F:133-363/Product: capsid antigen core protein p24CA #status predicted <P24>

F:364-377/Product: core protein p2 #status predicted <CP2>

F:378-432/Product: nucleocapsid core protein p1NC #status predicted <CP7>

F:392-405/Region: zinc finger CCHC motif

F:413-426/Region: zinc finger CCHC motif

F:433-448/Product: core protein p1 #status predicted <CP1>

F:449-500/Product: core protein p6 #status predicted <CP6>

F:2/Modified site: myristylated amino end (Gly) (in mature form) #status predicted

F:392-395, 400-405/Binding site: zinc (Cys, Cys, His, Cys) #status experimental

F:413, 416, 421, 426/Binding site: zinc (Cys, Cys, His, Cys) #status experimental

Query Match

Best Local Similarity 77.3%; Score 34; DB 1; Length 500;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

DB 454 EPTAPPEE 461

QY 1 KPSSPPEE 8

DB 454 EPTAPPEE 461

## RESULT 48

gag polypeptide - human immunodeficiency virus type 1 (isolate CDC-451)

N:Alternate names: core polypeptide

N:Contains: core protein p15; core protein p17; core protein p24

C:Species: human immunodeficiency virus type 1, HIV-1

C:Date: 30-Jun-1988 #sequence\_rev1501 30-Jun-1988 #text\_change 16-Jul-1999

C:Accession: A25523

R:Dees, S.M.; Kalyanaraman, V.S.; Casey, J.M.; Striyvasan, A.; Andersen, P.R.; Devare, Proc. Natl. Acad. Sci. U.S.A. 83, 8380-8384, 1986

A:Title: Molecular cloning and primary nucleotide sequence analysis of a distinct human

A:Reference number: A94136; PMID:87041461; PMID:3490666

A:Accession: A25523

A:Molecule type: DNA

A:Residues: 1-500 <DESS>

A:Cross-references: GB:M13136; NID:G326459; PIDN:AAA44306.1; PID:G326462

C:Genetics:

A:Gene: gag

C:Superfamily: AIDS-related virus gag polypeptide

C:Keywords: core protein; polypeptide

F:1-132/Product: core protein p17 #status predicted <P17>

F:133-391/Product: core protein p24 #status predicted <P24>

F:392-500/Product: core protein p15 #status predicted <P15>

Query Match 77.3%; Score 34; DB 1; Length 500;

Best Local Similarity 62.5%; Pred. No. 3.7e+02; Indels 0; Gaps 0;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8

DB 454 EPTAPPEE 461

DB 454 EPTAPPEE 461

## RESULT 49

gag polypeptide - human immunodeficiency virus type 1

C:Species: human immunodeficiency virus type 1, HIV-1

C:Date: 19-Mar-1997 #sequence\_rev1501 19-Mar-1997 #text\_change 26-Aug-1999

C:Accession: S33979

R:Cartini, F.

submitted to the EMBL Data Library, November 1991

A:Reference number: S33979

A:Accession: S33979

A:Molecule type: mRNA

A:Residues: 1-500 <CAR>

A:Cross-references: EMBL:Z11530; NID:G60192; PIDN:CAA77621.1; PID:G60193

C:Superfamily: AIDS-related virus gag polypeptide

Query Match 77.3%; Score 34; DB 2; Length 500;

Best Local Similarity 62.5%; Pred. No. 3.7e+02; Indels 0; Gaps 0;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8

DB 454 EPTAPPEE 461

## RESULT 50

gag polypeptide - human immunodeficiency virus type 1 (isolate ARV-2)

N:Alternate names: core polypeptide

N:Contains: core protein p15; core protein p17; core protein p24

C:Species: human immunodeficiency virus type 1, HIV-1

A:Note: host Homo sapiens (man)

C:Date: 17-May-1985 #sequence\_rev1501 17-May-1985 #text\_change 16-Jul-1999

C:Accession: A03947

R:Sanchez-Pescador, R.; Power, M.D.; Barr, P.J.; Steimer, K.S.; Stemple, M.M.; Brown-Shi

Science 227, 484-492, 1985

A:Title: Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2).

A:Reference number: A04003; PMID:85090453; PMID:2578227

A:Accession: A03947

A:Molecule type: DNA

A:Residues: 1-502 <SAN>

A:Cross-references: GB:K02007; NID:G328658; PIDN:AA59875.1; PID:G328661

C:Comment: Cleavage sites that yield the mature core proteins remain to be determined.

C:Genetics:

A:Gene: gag

C:Superfamily: AIDS-related virus gag polypeptide

C:Keywords: AIDS; core protein; immunodeficiency; polypeptide

F:1-134/Product: core protein p17 #status predicted <P17>

F:135-393/Product: core protein p24 #status predicted <P24>

F:394-502/Product: core protein p15 #status predicted <P15>

Query Match 77.3%; Score 34; DB 1; Length 502;

Best Local Similarity 62.5%; Pred. No. 3.7e+02; Indels 0; Gaps 0;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSSPPEE 8

DB 456 EPTAPPEE 463

Search completed: June 18, 2004, 13:00:26  
Job time: 29 secs

**This Page Blank (uspto)**